# AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 14, NUMBER 6 NOVEMBER–DECEMBER, 1946

On Micrometeorology*

H. K. SCHILLING, C. E. DRUMHELLER, W. L. NYBORG AND H. A. THORPE
The Pennsylvania State College, State College, Pennsylvania

THIS is a story of adventure in research experienced by a group of physicists whose work during the war led them, almost forcibly, out of their fields of special interest and thereby revealed what was for them a new world. Their excuse for telling the story is that they suspect many of their professional colleagues—ordinary physicists like themselves, with the same sort of training and lopsided reading habits—of being as oblivious to the existence of this interesting world as were they, and also as susceptible to its charms and fascinations. Moreover, they entertain the hope that their account may entice others, both teachers and students, to do some investigating of their own in this field, for, after all, the amount of work done in it to date is pitifully small.

WHEREOF WE WRITE

It is micrometeorology of which we write, the science of the weather in very small regions of space and very short intervals of time.[1] Like meteorology—which, from our point of view, might well be called "macrometeorology"—it is concerned, among other things, with temperature, wind velocity, humidity, and barometric pressure. In a sense, it is the ensemble of their values at a given time and place that we speak of as the weather. Meteorology concerns itself—at least, so it seems to us—with average values of these quantities, or with values which, though they may be measured at a point, are nevertheless considered to be more or less representative of comparatively large geographic areas, such as cities, or counties, or still larger regions. Moreover, the "weather man" measures quantities averaged over fairly large time intervals. He deliberately uses "sluggish" instruments which ignore rapid variations in time and thus yield the desired averages.

Unlike meteorology, micrometeorology deals with local and instantaneous values, and with the departures from the means. Thus it seeks to know, for instance, what the temperature is "here" and "now" and what its derivatives are with respect to both space and time.

We had the good fortune to carry on micrometric studies not only under temperate zone conditions, in Pennsylvania, but also in the tropics, in Panama. Here we frequently worked at a large airport, where there was a weather station with a staff of well-trained meteorologists and weather observers. These men were very cooperative and made all their data available to us. They could give us the temperature, say,

* Part of the data presented in this paper was obtained under contract with OSRD, Contract OEMsr-1210, SC-105.

[1] Since the term "micrometeorology" has probably not yet become standarized, this definition may be as good as any other.

for any time of any day—for the field as a whole, which was several square miles in extent. However, when we asked how the temperature above the hard surface runway differed from that over the grass beside it or how the wind velocity 10 ft above the ground compared with that at 2 ft, they could not help us. Clearly, such questions were outside their field, and related to the microweather at the port; and certainly they were not being paid to interest themselves in such matters, or to gather data concerning them.

A BIT OF HISTORY

Micrometeorology, or microclimatology, actually is not a brand new science. A not inconsiderable amount of work has been done on it in Germany,[2] especially since 1925, though some dates back nearly to the beginning of the century. Some work also has been done in Great Britain, and a very little in America. With the advent of the war the subject took on added significance and, as the result of the development of improved technics, underwent considerable progress.

The original interest in this subject seems to have arisen from practical problems in horticulture, especially viniculture, where it succeeded in shedding light upon the possibilities of frost control, of air drainage, and so forth. More recently a knowledge and understanding of small scale weather phenomena have been basic to the control of smoke and fog screens, in developing chemical warfare technics, and in studies of the transmission of various types of waves through the lower troposphere.

HOW WE GOT INTO IT

Our own interest came out of an investigation of the propagation of ultrasonic waves outdoors.[3] We had been asked to find out how far it might be possible to send and receive ultrasonic signals over various types of terrain, including dense tropical jungles, in all kinds of weather, day and night. Also we were to ascertain how reliable such signaling might be under various conditions. We soon discovered that the phenomena of atmospheric transmission were extraordinarily complex and that it was disconcertingly difficult to duplicate results from day to day, or place to place. On a given day we usually obtained vastly different results over grass than over, say, bare ground. Moreover, a signaling range at an elevation of 10 ft usually differed from that obtained within 1 ft of the ground. Nor was transmission at night like that during the day. Hence we were forced to recognize that the atmosphere might not be the simple, uniform medium it is often considered to be.

It is well known, of course, that the velocity of sound is determined in part by temperature, and that the refraction of sound depends partly upon temperature gradients in the propagating medium. Our curious ultrasonic findings led us to suspect, therefore, that the temperature within about 12 ft of the ground, where all our transmission experiments were made, might often vary with elevation much more rapidly than we had imagined formerly, that it probably depended to a surprising extent upon the surface covering of the ground, and that the difference between the temperature situation at night and during the day might involve much more than was apparent at first glance. Moreover, there were similar reasons to suggest that in all likelihood the motion of the air near the ground is also a much more complicated phenomenon than ordinary experience might indicate.[4]

It became apparent, therefore, that if we were to understand atmospheric ultrasonics it would be necessary to acquaint ourselves with fundamental micrometeorological phenomena, and that if ultrasonic measurements at a specific time and place were to be comprehensible at all, they would have to be accompanied by simultaneous measurements of the state of that particular part of the atmosphere transmitting the sound.[5]

All the data presented here are our own. Often they are typical of what others have

[2] An excellent book on this subject, the best we know of, is Rudolf Geiger's *Mikroklima und Pflanzenklima*, which constitutes vol. 1, part *D*, of the *Handbuch der Klimatologie* (Bebrüder Borntraeger, Berlin, 1930). A rather complete bibliography is included. It would be difficult to find any scientific treatise that makes more interesting reading.

[3] Our findings on atmospheric ultrasonics have been reported before the Acoustical Society of America and will be published in the journal of that society.

[4] Vern O. Knudsen has made some pertinent and stimulating remarks on the inhomogeneity of the atmosphere in an article entitled "The physicist in the new world," *Am. J. Physics* 11, 74 (1943).

[5] According to the physicist's definition, ultrasonic waves transmit "sound" even though it may be inaudible to a normal human ear.

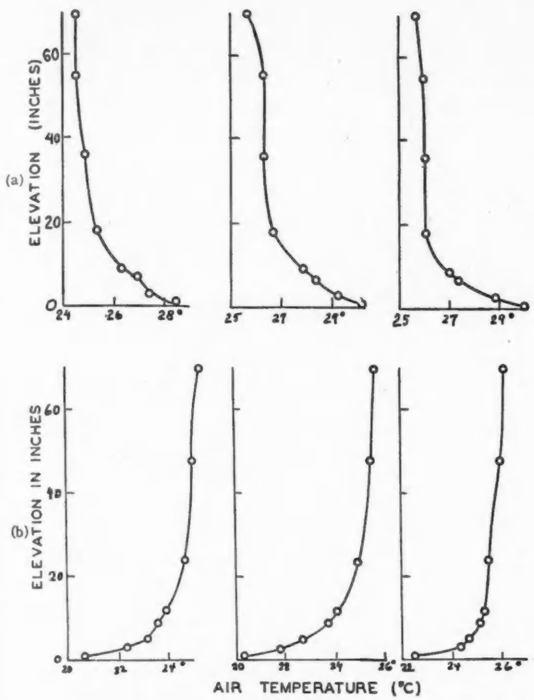


FIG. 1. Typical temperature profiles at the ground: (a) lapses over macadamized road; (b) inversions over short grass.

found and we do not claim them to be new. On the other hand, some of our most interesting findings are apparently new, for, so far as we are aware, they have not been reported before in print.

THE TEMPERATURES OF THE AIR

At Different Altitudes

Practically speaking, the most pressing problem was to learn how the temperature varies with altitude near the ground. Our first attack upon it was with ordinary laboratory mercury thermometers. A 10-ft pole, with a series of parallel transverse holes through it, was stuck in the ground. The thermometers were supported in the holes, and, when the sun shone, the pole was oriented in such a way that the thermometer bulbs were in the shadow of the pole. Later, supports of better design were constructed. Highly polished metal shields served to protect the thermometers from radiation by the sun and

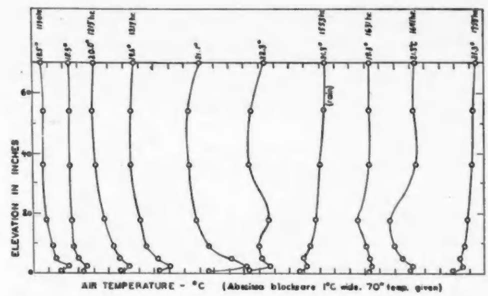


FIG. 2. Series of successive temperature profiles obtained during one afternoon.

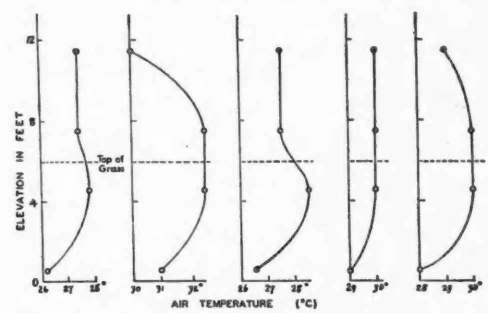


FIG. 3. Profiles, taken at 12-min intervals, depicting temperatures in and above tall, dense tropical grass.

the ground, and were so designed that the air around the thermometers would not become stagnant so long as there was any wind at all.

Typical results are illustrated in Fig. 1. These curves, called "temperature profiles," show that at times the temperature varies astonishingly with elevation. When the temperature decreases with increasing height, as in (a), there is said to be a temperature lapse; an increase with elevation, as in (b), is called a temperature inversion.

It should be noted that the gradients in evidence here are tremendous compared with those ordinarily encountered in meteorology. As is well known, the adiabatic lapse rate is approximately 1°C per 100 m. Usually at some distance above the earth the lapse rate is even less, approximately 1°C in several hundred meters. In Fig. 1, however, there appear over-all temperature differences of about 5°C, and temperature gradients of several degrees per foot within 1 ft of the ground. These data were obtained in May in Pennsylvania. In midsummer it is not uncommon to find still steeper gradients, both in

Pennsylvania and in Panama. A temperature difference of 10°C between heights of 1 in. and 6 ft is not at all rare. The reason for this is that the ground is often 20°C warmer than the air a few inches above it.

Effect of Kind of Ground

Figure 1 also presents evidence that the nature of the ground surface, or covering, is potent in determining temperature relations in the air within a few feet of it. All the curves of the figure were obtained during the same afternoon; the lapses occurred above a macadamized road, and the inversions over short grass in a pasture, less than a quarter mile away. The weather, from the viewpoint of macrometeorology, was the same for both locations. One indication of this is the fact that the temperatures at an altitude of 6 ft were essentially the same for both locations. (See Fig. 1.)

Nor is it necessary to choose locations as far apart as a quarter mile to illustrate this point. It is common to find decidedly different temperature gradients at points 20 ft apart over different surfaces.

For ultrasonics, as well as other sciences in which weather conditions may play a critical role, this is exceedingly important. It means that the experimenter must be wary lest he be deceived by inappropriate data. Thus it may be quite incorrect in some types of computations to use the temperature reported by a weather station, even if it is very close to the site of transmission experiments, because that temperature may have been taken too far above the ground. It must be remembered that the temperature at an elevation of, say, 6 ft may well be the same over a countryside several square miles in area, while at an elevation of only 1 or 2 ft, different microweather may be encountered every time one passes from one field to another, or comes to a highway, river or pond.

The underlying causes of this ground effect are not difficult to imagine. Without going into details, it is clear that the temperature gradients in the bottom stratum of air must depend, among other factors, upon the temperature of the ground. This, in turn, depends upon the thermal conductivity of the ground, upon its radiation absorptivity, reflectivity and emissivity. Soil texture and color are potent, partly, of course, because they affect the other quantities. The presence of vegetation and the topography of the terrain also are important factors, as will be seen in connection with the next topic.

Stratification

Figure 2 presents a series of temperature profiles obtained at irregular intervals during one afternoon at a site in Pennsylvania. They reveal several important facts. First, circumstances may be such as to produce definite temperature stratification of the air. Thus all profiles in Fig. 2 have a pronounced, though spatially small, inversion extending up from the ground 2 or 3 in. This means that there was a thin stratum of air in which the temperature situation was uniquely different from that in the higher strata. This was caused by grass, about 3 or 4 in. high, which shaded the ground. The air at the top of the grass was warmer than that below it.

Stratification on a larger scale is depicted by Fig. 3. This time only four thermometers were used, at elevations of 0.5, 4.5, 7.5 and 11.5 ft. The lower two were in dense, tropical grass 6 ft tall. The five profiles were taken at approximate 12-min intervals, between 10 and 11 o'clock one morning in Panama. As might be expected, an inversion existed in the grass to within about a foot of the top, throughout the hour of observation. Above the grass the gradient was either a lapse or zero.

Stratification in the two cases just considered was caused by the vegetation covering the ground. The topography of the ground may also cause contiguous layers of the atmosphere to differ in temperature, as well as in various other ways. Thus, hills or hillocks, valleys or slight depressions, embankments and terraces all have their effect in determining the microweather in their neighborhood, either directly or indirectly.

Variation of Profiles with Time

As noted earlier, we found it difficult to get the same results at different times over the same terrain. Micrometeorologically, this is now quite understandable. Both Figs. 2 and 3 provide typical evidence that the temperature structure of the air within a few feet of the ground often

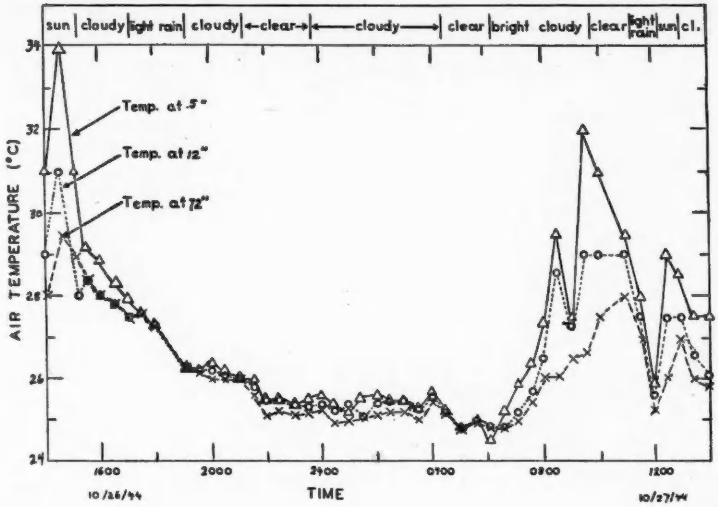


FIG. 4. Twenty-four hour cycle of temperatures over short grass in Panama, rainy season, at three heights: 0.5, 12 and 72 in. Accompanying over-all weather conditions also are indicated.

changes considerably with time. During the hour's observation yielding Fig. 3 the temperature varied continually both inside and outside the grass. Thus at 6 in. above the ground the successive readings were approximately 26°, 31°, 26.5°, 29°, 28°C. The changes at other elevations were similarly rapid and large.

One frequent cause of such changes is the passage of clouds overhead. It is interesting that the strong inversion in the grass persisted throughout these variations, and that the variation in the inversion gradient was slight compared to that of the temperatures itself. Another important fact strikingly revealed by Fig. 2 is that the temperature does not by any means always vary monotonically with elevation, as was the case in Fig. 1. At times the situation is complex indeed; lapses and inversions alternate both in space and in time.

Diurnal Cycles

It is well known that sound transmission is often much better at night than during the day. We found this to be true also in ultrasonics. Ranges were longer and intensities were steadier. Just why this was so was not at all obvious, despite the ordinary textbook explanations. Therefore, to identify underlying causes, and to discover the micrometeorological difference between day and night, we decided to make continuous observations for several days at a time.

We found that in Panama during the rainy season there are rather definite diurnal micrometeorological cycles—just as there are meteorological cycles—which recur with surprising regularity. Figure 4 shows such a cycle for an open, unforested terrain. The triangles symbolize the temperatures for an elevation of only 0.5 in., the circles for 12 in. and the crosses for 72 in. The triangles are predominantly above the circles, and these in turn above the crosses, indicating that temperature lapses predominated. An inversion occurred only once, namely, at six o'clock in the morning, about sunrise. The temperature differences, and therefore the gradients, were largest when it was sunny and clear. During the night they became negligible.

Clearly then, as far as temperature effects are concerned, nighttime in Panama is indeed very different from daytime. Not only is it cooler then, but the temperature is more uniform. Moreover, because the temperature differences are smaller at night, the air is more stable—and that is of great significance, as will appear later.

Effect of Rain

It will be observed that during the 24-hr period depicted by Fig. 4 it rained only twice, once at sunset (1800 hr), and again at noon the next day. Both times the temperature differences for various heights were reduced to approxi-

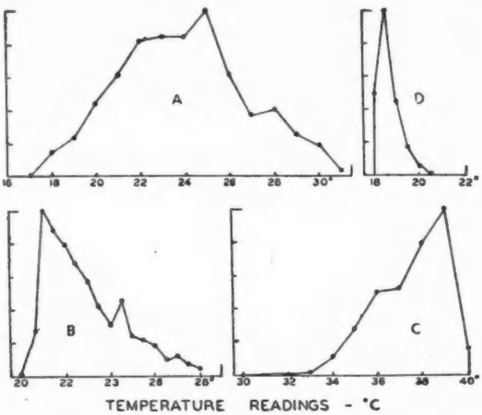


FIG. 5. Various frequency distributions of temperatures, observed under different circumstances with sensitive resistance thermometer, at a point, at intervals of 6 sec, for half-hour periods.

mately zero. Each time, too, temperature differences were reestablished almost immediately after the rain. Whether this is true in general we do not know. Our data during rain are not extensive enough to justify broad claims. However, we do know from repeated observations that there is such an effect at times, and the micrometeorological data make sense when correlated with simultaneous ultrasonic data.

"Instantaneous" Temperatures

A worker first entering the field of atmospheric acoustics is usually surprised by an extremely interesting and puzzling phenomenon, namely, the erratic and unpredictable variation of intensity.[6] Even when the output of the source itself is perfectly steady, the signal intensity always fluctuates out of doors, sometimes as much as ±15 db, or even more, and several times per second. This seems to suggest that there are correspondingly rapid variations of the "microcondition" of the air, and, more particularly, that the temperature of the air at a given point may fluctuate much more rapidly than is indicated, say, by Fig. 2. After all, the mercury thermometers yielding those results were not particularly rapid in their response, and therefore could give only average temperatures corresponding to fairly long intervals of time.

What would happen if more sensitive devices were used, say resistance thermometers employing short, thin wires of small thermal capacity? To find out, we built several such instruments using nickel wires 0.0005 in. in diameter, 4 cm long, wound in tiny helixes of about 42 turns. The resistance was measured by a Wheatstone bridge with a portable, sensitive, rapid galvanometer, calibrated to read temperature directly.

Our hunch was correct. At times the temperature at a point does fluctuate rapidly indeed, sometimes 10°C or more, within a second or two. These fluctuations, like those of acoustic and ultrasonic intensity, are erratic and unpredictable, often occurring several times per second.[7] Figure 5 presents typical data. Each graph is for one point and shows the relative number of times various temperatures were observed there during a period of about 30-min intervals. The readings were taken at 6-sec intervals, and each is an estimate of the average temperature for about a half second. For Fig. 5, *A* these virtually instantaneous temperatures ranged from 17° to 31°C, though the reading of a mercury thermometer, within a few inches of the nickel wire, remained essentially constant. The vertical scale is in arbitrary units, the most frequently occurring temperature having the largest ordinate. The relative frequencies of appearance of other

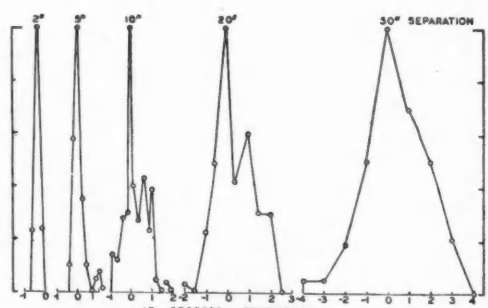


FIG. 6. Frequency distributions of temperature differences for increasing thermometer separations.

[6] In regard to this phenomenon also, Professor Knudsen makes interesting observations; see reference 4.

[7] This was first discovered, so far as we are aware, by a group of workers under the direction of R. E. Wegel, working under OSRD contract at Duke University. We are indebted to them for this information, which apparently they chose not to publish in their OSRD reports or elsewhere.

temperatures are plotted as percentages of the maximum ordinate. For graph *A* the modal temperature, the one observed most often, was 25°C. Graph *B*, representing 300 readings taken in the same way on another day, covers a range of 8°C, from 20° to 28°C. The mode was 21°C; that is, the distribution was definitely asymmetric. It should be emphasized that these temperatures are thoroughly unpredictable as to their sequence in time. Thus, under conditions represented by graph *A*, a typical succession of readings might well be 20°, 28°, 30°, 18°, 30°, 26°, 25°, 24°, 19°, 29°C, ...

Discrete Air Packets

It seems reasonable to infer from these findings that the atmosphere consists of more or less discrete bodies of air with different temperatures, and that the temperature fluctuations are caused by their motion in space. On this hypothesis a distribution such as *A* would mean that the temperatures of such bodies are nearly evenly distributed about the average. On the other hand, according to graph *B* one would imagine the great mass of air to have a temperature very close to the mean, and that floating about in it are pockets predominantly warmer. Graph *C* would suggest that the "foreign bodies" were predominantly colder than their surroundings, and *D* that the air was practically homogeneous as far as temperature was concerned, when the data were taken.

If such discrete bodies of air do indeed exist in the lower atmosphere, how large are they? How can their diameters be measured? One method that suggests itself is to set up two sensitive thermometers very close together and observe how the difference of their readings varies as the thermometers are separated gradually. Conceivably, if they were close together, say only a few millimeters apart, they rarely would indicate a significant temperature difference, because both would usually be in the same temperature body. On the other hand, if they were a considerable distance apart, they would seldom be inside the same temperature volume simultaneously, and would therefore indicate large, erratically varying temperature differences. To test this hypothesis, we used two sensitive resistance thermometer wires as balancing

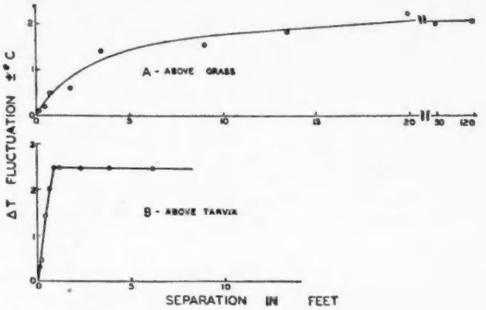


FIG. 7. Variation of range of temperature differences with increasing thermometer separations: *A*, above grass; *B*, above macadam.

arms of a Wheatstone bridge, with the galvanometer scale calibrated to indicate directly the temperature differences causing unbalance.

With this instrument we obtained the distribution curves of Fig. 6, all at the same location, on the same afternoon, when the weather remained essentially unchanged. It should be noted that the abscissas in this case refer to temperature differences, not temperatures. We see that, upon this occasion at least, the temperature differences did increase with the separation of the thermometers. For a separation of 2 in. the temperature difference was always less than 1°C, whereas for 50 in. the range of differences was approximately ±4°C.

Now, if the size of the air "blobs" under discussion has a "statistical" upper limit under a particular set of conditions, the observable temperature differences presumably would not increase indefinitely with separation, but would reach an upper limit and thereafter remain constant. Therefore to obtain hints on the probable dimensions of the packets it seemed desirable to plot the temperature difference as a function of separation. This is done in Fig. 7. The separations, which in this case were horizontal, appear as abscissas. The ordinates represent the 90th percentile temperature differences for corresponding separations, that is, in each case that value of the difference which exceeds 90 percent of all observed values. In the case of graph *A* the temperature difference reached a maximum when the separation was about 200 in. By contrast, in graph *B* the largest difference was attained when the separation was only 10 in.

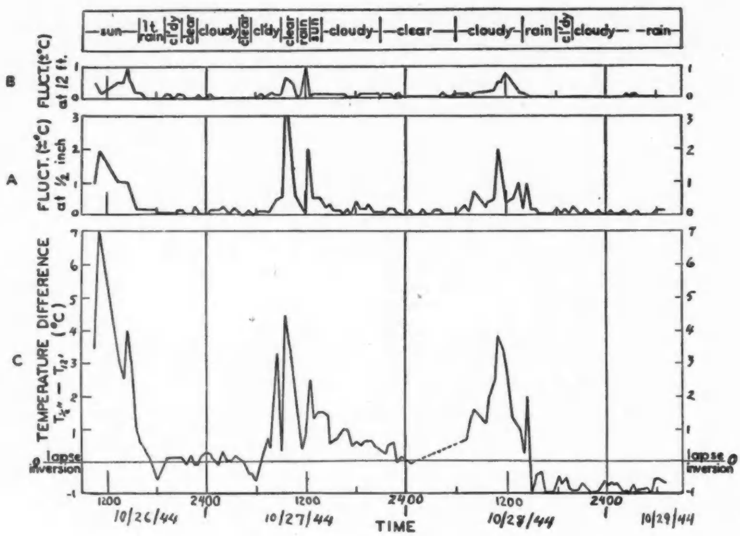


FIG. 8. Cyclic, diurnal correlation of vertical temperature differences with temperature fluctuations at two elevations: *A*, fluctuations at 0.5 in. above ground; *B*, fluctuations at 12 ft; *C*, differences of temperatures at heights of 0.5 in. and 12 ft *vs.* time.

Clearly such evidence is not conclusive. However, it is rather suggestive and tempts one to conclude that the two curves of Fig. 7 symbolize two vastly different structures or states of the air. In one case the diameters of air "bodies," in the direction of separation of thermometer elements, were probably about 15 ft, and in the other only 1 ft.

It is to be expected, of course, that the size of such inhomogeneities in the air, as inferred from studies of temperature fluctuations such as we have described briefly, should vary with the time of day, elevation, vertical temperature gradients, wind velocity, kind of ground surface, and probably other factors.

Figure 7 illustrates one such effect, that of the ground. The data were obtained on the same afternoon at locations not more than 100 ft apart, both at an elevation of 1 ft, *A* over a grass lawn, and *B* over a macadamized parking lot. Certainly, on that particular day the structure of the air within a foot of the ground was determined largely by the type of ground covering, relatively large "blobs" of air forming over grass, and much smaller ones over macadam.

The altitude effect is illustrated by Fig. 5, *C* and *D*. The former depicts temperatures observed at a height of only 2 in., while the latter is for a height of 6 ft—at the same location, the same afternoon. The conclusion seems to be that as the packets of different temperatures move upward they tend to coalesce and form a more homogeneous mass of air, as might be expected theoretically.

That there is a time effect is shown by the data of Fig. 8, obtained in Panama from continuous observation for three days over short grass. In *A*, estimated mean temperature fluctuations are plotted as a function of time. The measurements were made with a resistance thermometer placed 0.5 in. from the ground. A diurnal cycle, which repeats itself in a rather predictable way, is clearly in evidence. Fluctuations are largest during the day, from 6 A.M. to 6 P.M., and are almost negligible at night. The reason for this is suggested by graph *C*, where the difference of temperatures at elevations of 0.5 in. and 12 ft is plotted against time. When the curve lies above the zero axis there is a temperature lapse, when below an inversion. Curves *A* and *C* together show that when the temperature difference representing a lapse is largest the temperature fluctuations are greatest. When the temperature difference is negligible, the fluctuations are negligible.

All this is just what we would expect theoretically. When the ground is hot and the lowest layer of air is hotter and less dense than that

above it, the air is unstable mechanically. Under those conditions chunks of hot air should break loose from the ground and float upward while cooler air falls to replace them. The temperature at any point near the ground would therefore fluctuate violently as the unequally heated air packets move across it. Obviously such an effect should be maximum when the sun is shining, and minimum, or even nonexistent, at night.

Figure 8, *B* illustrates further the altitude effect referred to previously. Here the fluctuations observed at a height of 12 ft are plotted. The same diurnal cycle is apparent; however, as expected from earlier considerations, the fluctuations are smaller in amplitude, because the air is more homogeneous at this level.

General Conclusion

All these findings lead to the conclusion that, in regard to temperature, the atmosphere within a few feet of the ground is a surprisingly complex and variable structure. It is therefore virtually meaningless to talk of *the temperature* of the air.

MOTIONS OF THE AIR

Since ultrasonic waves in the atmosphere are affected by wind quite as much as by temperature, it was necessary for us to investigate the motions of air also with micrometeorological technics.

Hot-wire anemometers were used for the measurement of wind velocity. The sensitive elements were platinum wires about 1 cm long and 0.0004 in. in diameter. These instruments were very rapid in their response and were capable of indicating as many as ten changes of speed per second. They were essentially nondirectional in a plane normal to a wire and were for most experiments mounted in such a way as to be nondirectional horizontally.

Wind phenomena are even more complicated than those of temperature, one reason being that wind velocity involves direction as well as magnitude. Nonetheless, many relationships parallel those for temperature, and so it will not be necessary here to give as many details as in the preceding section.

To begin with, usually there are steep gradients of wind at the ground. This is illustrated by the typical wind-speed profiles of Fig. 9 obtained in Panama in unforested regions. As might be expected, the velocity within an inch or two of the ground is often nearly zero, though this is by no means always true.

In Fig. 10 the dots represent average wind velocities. The two curves indicate the variation of the mean velocities with time at elevations of 1 in. and 12 ft, respectively. They illustrate the fact that in Panama these velocities and velocity gradients follow a diurnal cycle similar to that of daily temperature variation. Usually the wind velocity, and gradients, are greater during the day than at night.

Our rapidly responding anemometers revealed the fact, apparently not commonly known, that wind velocity, though it may appear constant when observed with a rotating-vane anemometer, usually fluctuates quite as erratically and rapidly as we have found the temperature to fluctuate. The vertical bars in Fig. 10 indicate the estimated mean amplitudes of fluctuation, for short time intervals. It will be noticed that during that particular 24-hr period the amplitudes frequently were ± 25 percent of the mean velocity. Moreover, in open, unforested areas the fluctuations vary cyclically, just as temperature fluctuations do, being most pronounced in the day and much less so at night.

The question arises here, just as it did with regard to temperature, whether the wind fluctuations vary from point to point at a given time. To answer this we devised a speed-difference meter, which indicated directly the difference between the speeds measured by two sensitive

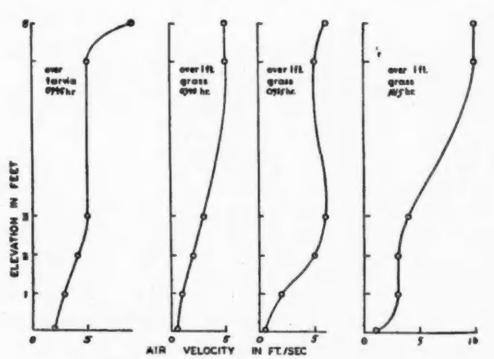


FIG. 9. Typical wind profiles in open, unforested terrain.

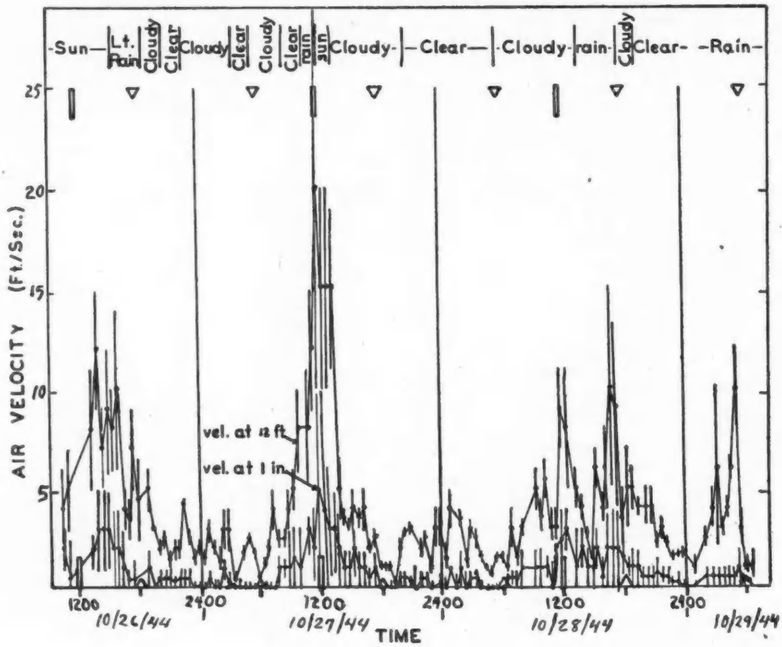


FIG. 10. Cyclic, diurnal variation of wind and wind fluctuations at two elevations.

elements whose separation could be varied at will. Experiments with this instrument are analogous to those represented by Figs. 6 and 7. They showed somewhat similar results, namely, that with respect to the wind also the air is sometimes "grainy"; that is, the speed difference is almost never zero even when the two points of observation are only a few feet apart.

Thus far we have discussed only the magnitude of the wind velocity. The problem of measuring the direction, or definitely isolating its effects, is much more demanding and requires rather elaborate experimental technics, the discussion of which would lead us too far afield. However, anyone interested in the matter can convince himself by simple methods that at any given point the direction of motion of the air often varies quite unpredictably and rapidly, and that, furthermore, it varies greatly from point to point. This is, of course, merely another way of saying that near the ground the air is often exceedingly turbulent, certainly much more so than the authors had suspected before they looked into the matter themselves.

One simple way to observe this is to use arrays of tiny wind vanes,[8] rotating with exceedingly small friction about three different axes that are mutually perpendicular. Such an array often shows that, at a given time, the wind may be eastward at one point, westward at another point just a few feet away, and upward at still another nearby point. A still simpler, but very revealing, method is to use small soap bubbles. With the inexpensive but effective wire loop bubble outfits commonly available now in ten-cent stores it is possible to make several, sometimes as many as a dozen, bubbles in quick succession, in a fraction of second. Such bubbles, made with standard commercial soap solution, are remarkably tough and long lived. What they can be made to show about the motion of the air is astonishing. On some days a group of bubbles initially occupying a volume of several cubic feet is quickly dispersed, some going up, others down; some moving rapidly, others slowly; some moving almost rectilinearly for considerable distances, while others are obviously caught in

[8] Ours were made of stiff paper fastened transversely to tiny glass tubes slipped over and free to turn about fine wires used as supports.

vortices. It is revealing, also, to make observations on different days and to classify days in terms of the behavior of the bubbles, that is, in terms of the kind and amount of turbulence, or variability and "graininess" of the wind.

HUMIDITY

In atmospheric ultrasonics water vapor plays an important role because it is a potent factor determining absorption of sound by the air. Hence humidity directly affects signal intensity. It would seem highly desirable, therefore, to discover how water vapor is distributed in the air.

Unfortunately, from the viewpoint of instrumentation this is extremely difficult, because, to the best of our knowledge, there exists now no instrument capable of measuring either relative or absolute humidity during a very short time in a minute volume, that is to say, instantaneously at a point. Nonetheless it is possible to show with a set of hygrothermographs that the distribution of water vapor usually is not uniform.

At times there are vertical gradients of water vapor content near the ground. This can be shown with such instruments because they simultaneously record temperature and relative humidity, from which the absolute humidity may be computed. One can show that the absolute humidity over plowed ground is often different from that over a meadow of green grass, or over a macadamized parking lot. Likewise on a lawn the water content of the air is not always the same over a patch of dry grass as it is over a spot of healthy green grass. When such inhomogeneities exist, a wind should blow them about; therefore, within a few feet of the ground the absolute humidity should fluctuate. However, so far as we are aware, this has not yet been demonstrated experimentally; the development of a method of doing it is one of the important, challenging problems of micrometeorology awaiting solution.

Thus ends the writing of our story. May its reading be as great a pleasure as has been its telling. The intention has been to give, not a systematic review, but an introduction by the "sampler" method—enough sampling, we hope, to show that micrometeorology is a field worthy of the attention and efforts of the most highly skilled researcher and most completely equipped laboratory, and yet entirely suitable for investigations requiring only humble resources. It is replete with opportunities for valuable, original work even on the undergraduate level.

We are greatly indebted to many persons for help during the progress of our work, particularly to Dr. I. Rudnick and Mr. Paul Oncley, at Duke University during the war, to Dr. Wayne Hales, then of Rutgers University, and to Dr. M. P. Givens and Mr. J. S. Saby, our colleagues who were with us in Panama as experts on electric equipment.

Placement Register at Annual Meeting of the Association

IN response to many requests from physicists and employers of physicists, the American Institute of Physics will operate another one of its Placement Registers in connection with the joint meeting of the Association and the American Physical Society on January 30–31 and February 1, 1947. The Register affords an effective supplement to the continuing Placement Service carried on by the Institute. The latter service provides an opportunity for those desiring new employment to place their records on file at the Institute office; here they may be consulted by employers or sent out to employers for consideration.

The Register itself provides an on-the-spot means for bringing applicants and employers' representatives together. One or more rooms will be reserved for the Register and interviews will be scheduled as requested. Because employers will have opportunity to consult the records of applicants before the interview period begins, pre-registration on the part of applicants is essential for the maximum effectiveness of the Register. Registration forms and further information may be obtained from the Institute, 57 East 55 Street, New York 22, New York.

General Equations of Electricity and Magnetism That Are Not Dependent on the Systems of Units Used in Making Computations

A. G. Worthing
University of Pittsburgh, Pittsburgh 13, Pennsylvania

PHYSICS textbooks show two radically different points of view relative to the meanings of the symbols used in equations. One, the "physical view," is that the symbols stand for physical quantities independent of any units or systems of units of measurement. The other, the "numeric view," is that the symbols stand for numerics and are thus inseparable from the units of measurement or system of units of measurement for which the equation is designed. For the simple case of a moving body and the equation $v=s/t$, the "physical view" regards v, s and t in turn as the average speed, the distance transversed and the time spent in traveling that distance, all of which are physical quantities. The "numeric view," however, to choose one of many possibilities, regards v, s and t as the number of miles per hour, the number of miles and the number of hours associated with the movement of the body, all of which are numerics.

The "physical view" recognizes concepts of multiplication and division that go beyond the elementary concepts which suffice for the "numeric view." To illustrate, the elementary concept of division does not permit the division of 120 mi by 3 hr. However, there is an understandable physical concept of speed to correspond to the 40 mi/hr which we obtain, in accord with the "physical view," by treating 120 mi and 3 hr as we treat algebraic quantities in division. The treatment of the incommensurable units *mile* and *hour* is, of course, just that given the algebraic unknowns x, y, $\cdots$ in the corresponding algebraic solutions. So far as we are concerned, the justification for this extended "physical view" of division and, of course, for a similar view of multiplication lies in the fact that the procedures are simple and lead to understandable results.

Except for the fields of electricity and magnetism, there seem to be no fundamental difficulties associated with either procedure. The forms of equations for the two points of view, except for those fields, are identical as long as the equations for the "numeric view" are written for well-recognized systems of units, as is usually but not always done. Illustrating an exception, we have the common engineering equation for power,

$$P=\frac{1}{550}fv, \tag{0}$$

which includes only the numeric part of the conversion factor 1 hp/(550 ft lb/sec) and therefore can serve only those who hold the "numeric view."

In the specified fields, however, depending on the system of units, the "numeric view" frequently demands different equations for the same physical phenomena. Illustrations are the expressions associated with the magnetomotive force of a closed path linking a wire carrying an electric current. With $\mathbf{H}$, $\mathbf{s}$, I and n representing magnetic field strength, distance, electric current and number of linkages, the common electrostatic and electromagnetic and the mks systems demand

$$\oint \mathbf{H}\cdot d\mathbf{s}=4\pi nI, \tag{1}$$

while the Gaussian system requires

$$\oint \mathbf{H}\cdot d\mathbf{s}=(4\pi/c')nI, \tag{2}$$

of which c' is 3×10^{10}, the numeric part of c, the velocity of light in free space, when expressed in centimeters per second. The Heaviside-Lorentz and the MKS systems demand, respectively,

$$\oint \mathbf{H}\cdot d\mathbf{s}=(1/c')nI, \tag{3}$$

and

$$\oint \mathbf{H}\cdot d\mathbf{s}=nI, \tag{4}$$

while the hybrid em-MKS system of many

present-day textbooks requires

$$\oint \mathbf{H}\cdot d\mathbf{s} = (4\pi/10)nI. \tag{5}$$

The common electrostatic (st) and electromagnetic (ab) systems are well known. However, it may be appropriate to comment briefly on the other systems. The Gaussian (ga) system is identical with the common electrostatic system in electrical matters and the common electromagnetic system in magnetic matters. The Heaviside-Lorentz (hl) system (see Table I for the relative magnitudes of certain units of several systems), devised to eliminate the excessive occurrence of 4π in computations, differs from the Gaussian system through the selection of the hlcoulomb $[=1/\sqrt{(4\pi)}$ stcoul$]$ as its static unit of electric

TABLE I. Physical quantities, their symbols and equivalent values for their units. The symbol c' stands for the numeric part of the velocity of light in free space when expressed in centimeters per second; MKS-ϵ and MKS-μ are abbreviations for the MKS units of ϵ and μ. The values found in a single line are equal. The ratio of any two such equal values constitutes a conversion factor. For each system of units, the product of one unit of ϵ, one unit of μ and the square of one unit of velocity is unity. For each system of units, the product of the ϵ of free space, the μ of free space and the square of the velocity of light in free space is unity.

Physical quantity and symbol	Electrostatic (st)	Electromagnetic (ab)	Gaussian (ga)	Heaviside-Lorentz (hl)	mks	MKS
Distance, $\mathbf{s}$	1 cm	1 cm	1 cm	1 cm	10^{-2} m	10^{-2} m
Force, $\mathbf{f}$	1 dy	1 dy	1 dy	1 dy	10^{-5} newton	10^{-5} newton
Energy, U	1 erg	1 erg	1 erg	1 erg	10^{-7} j	10^{-7} j
Electric charge, Q	1 stcoul	$\frac{1}{c'}$ abcoul	1 stcoul	$(4\pi)^{\frac{1}{2}}$ hlcoul	$\frac{10}{c'}$ coul	$\frac{10}{c'}$ coul
Electric current, I	1 stamp	$\frac{1}{c'}$ abamp	1 stamp	$(4\pi)^{\frac{1}{2}}$ hlamp	$\frac{10}{c'}$ amp	$\frac{10}{c'}$ amp
Dielectric constant, ϵ	$1\frac{\text{stfar}}{\text{cm}}$	$\frac{1}{c'^2}\frac{\text{abfar}}{\text{cm}}$	$1\frac{\text{stfar}}{\text{cm}}$	$4\pi\frac{\text{hlfar}}{\text{cm}}$	$\frac{10^{11}}{c'^2}\frac{\text{farad}}{\text{m}}$	$\frac{10^{11}}{4\pi c'^2}$ MKS-ϵ*
Electric field strength, $\mathbf{E}$	$1\frac{\text{dy}}{\text{stcoul}}$	$c'\frac{\text{dy}}{\text{abcoul}}$	$1\frac{\text{dy}}{\text{stcoul}}$	$\frac{1}{(4\pi)^{\frac{1}{2}}}\frac{\text{dy}}{\text{hlcoul}}$	$\frac{c'}{10^6}\frac{\text{newton}}{\text{coul}}$	$\frac{c'}{10^6}\frac{\text{newton}}{\text{coul}}$
Potential difference, V	1 stv	c' abv	1 stv	$\frac{1}{(4\pi)^{\frac{1}{2}}}$ hlv	$\frac{c'}{10^8}$ v	$\frac{c'}{10^8}$ v
Capacitance, C	1 stfar	$\frac{1}{c'^2}$ abfar	1 stfar	4π hlfar	$\frac{10^9}{c'^2}$ farad	$\frac{10^9}{c'^2}$ farad
Pole strength, m	c' stpole	1 abpole	1 abpole	$(4\pi)^{\frac{1}{2}}$ hlpole	$\frac{1}{10^8}$ mkspole	$\frac{4\pi}{10^8}$ MKSpole
Permeability, μ	$\frac{1}{c'^2}\frac{\text{sthen}}{\text{cm}}$	$1\frac{\text{abhen}}{\text{cm}}$	$1\frac{\text{abhen}}{\text{cm}}$	$\frac{1}{4\pi c'^2}\frac{\text{hlhen}}{\text{cm}}$	$\frac{1}{10^7}\frac{\text{henry}}{\text{m}}$	$\frac{4\pi}{10^7}$ MKS-μ*
Magnetic field strength, $\mathbf{H}$	c' stoer	1 oersted	1 oersted	$\frac{1}{(4\pi)^{\frac{1}{2}}}$ hloer	10^3 mksoer*	$\frac{10^3}{4\pi}\frac{\text{amp turn}}{\text{m}}$
Magnetomotive force, $\mathfrak{F}$	c' stgil	1 gilbert	1 gilbert	$\frac{1}{(4\pi)^{\frac{1}{2}}}$ hlgil	10^3 mksgil*	$\frac{10^3}{4\pi}$ amp turn

* 1 MKS-$\epsilon = 4\pi\frac{\text{farad}}{\text{m}}$; 1 MKS-$\mu = \frac{1}{4\pi}\frac{\text{henry}}{\text{m}}$; 1 mksoer $= \frac{1}{4\pi}\frac{\text{amp turn}}{\text{m}}$; $\frac{1}{4\pi}$ mksgil $=1$ amp turn.

charge and the hlpole [$=1/\sqrt{(4\pi)}$ abpole] as its unit of magnetic pole strength. The MKS system is the common, so-called practical system in electromagnetism, but in the field of electrostatics it is the system that gives to the free-space dielectric constant, or permittivity, ϵ_0 a value of $(10^{11}/4\pi c'^2)$ MKS units of ϵ. (See note at end of paper.)

Another system, the mks, differs from the MKS system, which is to be largely used in this paper, in that it gives to ϵ_0 the value $10^{11}/c'^2$ far/m. The hybrid em-MKS system is composed of the magnetic part of the common electromagnetic system and the electric part of the MKS system. Adherents of the "numeric view" point to these differing equations as evidence for the impossibility of the "physical view."

In view of the various forms, Eqs. (1) to (5), required by the different "numeric views" to represent the same phenomenon, an obvious question arises. Are Eqs. (1) to (5) possibly on a par with Eq. (0) in omitting portions of conversion factors?

It is the purpose here to consider how, if at all, Eqs. (1) to (5), as a group, and other similar groups of equations in the fields of electricity and magnetism may be replaced by single equations. The basis for the procedure is the fundamental postulate of the "physical view," namely, *wherever a given set of physical conditions predicts a definite physical result, the phenomenon, if describable by an equation, should be describable by a single equation provided the symbols represent physical quantities.*

In a paper presented at the January, 1945 meeting of the American Association of Physics Teachers Ross[1] states:

> The laws of physics in their most general form are proportionalities. If they are written as equations, proportionality constants should be included. It is only in this way that they can be made independent of the system of units. . . . Some equations . . . are universally accepted definitions, in which case no proportionality constant is needed in the general form.

With the understanding that a proportionality constant may be the numeric 1, the present author is in complete agreement with Ross' statement, and has made use of it in developing a procedure.

Consider first the Coulomb equation of electrostatics, for which, according to the "numeric view," the common electrostatic, the common electromagnetic, the Gaussian and the common mks expressions are

$$f=\left[\frac{Q_1Q_2}{\epsilon s^2}\right]_{\text{es, em, ga, mks}}, \tag{6}$$

and the Heaviside-Lorentz and the MKS expressions are

$$f=\left[\frac{Q_1Q_2}{4\pi\epsilon s^2}\right]_{\text{hl, MKS}}. \tag{6a}$$

In place of these, from the "physical view," let us write

$$f=k\frac{Q_1Q_2}{\epsilon s^2}, \tag{7}$$

in which k, a proportionality constant, must be single valued. It may be a numeric. Rearrangement gives

$$k=\epsilon f s^2/Q_1Q_2. \tag{8}$$

Obviously, with the common electrostatic system of units in mind, k may be evaluated as

$$k=1\text{ st-}\epsilon\text{ dy cm}^2/(\text{stcoul})^2, \tag{9}$$

of which "st-ϵ" is to be read "the electrostatic unit of ϵ." Similar interpretations are to be given to ga-ϵ, st-μ, ga-μ, $\cdots$, which occur later.

Equation (9) is the form in which k would be expressed for computing purposes when ϵ, f, s and Q are given in common electrostatic units. Since the electric units of the Gaussian system are identical with those of the common electrostatic system, the k of Eq. (8) is also the convenient form when Gaussian units are involved. The convenient form for the common electromagnetic unit is

$$k=1\frac{\text{st-}\epsilon\text{ dy cm}^2}{(\text{stcoul})^2}\left(\frac{c'\text{ stcoul}}{\text{abcoul}}\right)^2\frac{\text{ab-}\epsilon}{c'^2\text{ st-}\epsilon}$$

$$=1\frac{\text{ab-}\epsilon\text{ dy cm}^2}{(\text{abcoul})^2}, \tag{10}$$

of which (c' stcoul/abcoul) and (ab-ϵ/c'^2 st-ϵ) are conversion factors and therefore equal to the

[1] D. Ross, *Am. J. Physics* **13**, 121 (1945). The complete paper, from which the quotation has been taken, has not been published.

numeric 1. The basic equalities underlying the conversion factors of Eq. (10) and of many equations that follow will be found in Table I. For additional data covering the field of electricity and magnetism, reference may well be made to W. R. Varner's monograph.[2]

The convenient forms for k for the Heaviside-Lorentz and the MKS systems are obtained similarly. They are

$$k=1\frac{\text{st-}\epsilon\text{ dy cm}^2}{(\text{stcoul})^2}\,\frac{1\text{ hl-}\epsilon}{1\text{ st-}\epsilon}\left(\frac{1\text{ stcoul}}{\sqrt{4\pi}\text{ hlcoul}}\right)^2=\frac{1}{4\pi}\,\frac{\text{hl-}\epsilon\text{ dy cm}^2}{(\text{hlcoul})^2} \tag{11}$$

and

$$k=1\frac{\text{st-}\epsilon\text{ dy cm}^2}{(\text{stcoul})^2}\,\frac{10^{11}\text{ MKS-}\epsilon}{4\pi c'^2\text{ st-}\epsilon}\left(\frac{c'\text{ stcoul}}{10\text{ coul}}\right)^2\frac{1\text{ n m}^2}{10^5\text{dy cm}^2}=\frac{1}{4\pi}\,\frac{\text{MKS-}\epsilon\text{ Newton m}^2}{\text{coul}^2}. \tag{12}$$

It is to be emphasized that, in accord with the "physical view," all of the values for k given in Eqs. (9) to (12) are equal. What is rather surprising is that the proportionality constant k may be viewed as a conversion factor which in Eq. (7), for instance, permits converting a unit for $(Q_1Q_2/\epsilon s')$ into units for f. Like all conversion factors[3] it will have the value of unity. Then Eq. (7) becomes

$$f=Q_1Q_2/\epsilon s^2. \tag{13}$$

When this is done, we must recognize that, as follows from Eq. (9) to (12),

$$1\text{ st-}\epsilon=1\text{ ga-}\epsilon=1\frac{(\text{stcoul})^2}{\text{dy cm}^2}=1\frac{\text{stfar}}{\text{cm}}, \tag{14}$$

$$1\text{ ab-}\epsilon=1\frac{(\text{abcoul})^2}{\text{dy cm}^2}=1\frac{\text{abfar}}{\text{cm}}, \tag{15}$$

$$1\text{ hl-}\epsilon=4\pi\frac{(\text{hlcoul})^2}{\text{dy cm}^2}=4\pi\frac{\text{hlfar}}{\text{cm}}, \tag{16}$$

$$1\text{ MKS-}\epsilon=4\pi\frac{(\text{coul})^2}{\text{Newton m}^2}=4\pi\frac{\text{far}}{\text{m}}. \tag{17}$$

Incidently, for ϵ_0, the ϵ of free space, we have

$$\epsilon_0=1\text{ st-}\epsilon=\frac{1}{c'^2}\text{ab-}\epsilon=1\text{ hl-}\epsilon=\frac{10^{11}}{4\pi c'^2}\text{MKS-}\epsilon. \tag{18}$$

For all purely electrostatic problems involving forces between static charges, Eq. (13) suffices for both representation and computation when cognizance is taken of Eqs. (14) to (17).

As a simple problem consider the repulsive force between two electric charges of 5 hlcoul each, 1 cm apart in a medium whose dielectric constant is 3 hl-ϵ. Substituting in Eq. (13), and using the definition of Eq. (16) for the hl-ϵ, we have

$$f=\frac{(5\text{ hlcoul})^2}{12\pi\dfrac{(\text{hlcoul})^2}{\text{dy cm}^2}4\text{ cm}^2}=\frac{25}{48\pi}\text{dy}. \tag{19}$$

Other relations in electrostatics cause little trouble once the various convenient forms for the k of the basic equation for the field have been established. Thus for the field strength E due to a charge Q_1, at a distance r in a medium characterized by ϵ, we have

$$E=f/Q_2=Q_1/\epsilon s^2. \tag{20}$$

For the static potential difference V between two points at distances s_1 and s_2 from charge Q_1 in a medium characterized by ϵ, whatever the system of units, we have

$$V=\frac{Q_1}{\epsilon}\left(\frac{1}{s_1}-\frac{1}{s_2}\right). \tag{21}$$

In the field of magnetics, except for the MKS unit, the treatment, starting with the coulomb expression for the force between two magnetic poles, is exactly the same as that given above. Corresponding to Eqs. (13) and (18) we should obtain

$$f=m_1m_2/\mu s^2, \tag{22}$$

$$1\text{ st-}\mu=1\frac{(\text{stpole})^2}{\text{dy cm}^2}=1\frac{\text{sthen}}{\text{cm}}, \tag{23}$$

[2] W. R. Varner, *The fourteen systems of units* (O.S.C. Cooperative Association, Corvallis, Ore., 1943). Those who use this excellent book should note one procedure which differs from that of the present author. For instance, Varner uses such an expression as (stcoul/abcoul), by itself without a preceding numeric, to represent the number of statcoulombs equivalent to 1 abcoul, whereas the present author treats the expression whenever it occurs as the ratio of the statcoulomb to the abcoulomb. Varner therefore equates (stcoul/abcoul) to c', whereas the present author equates it to $1/c'$.

[3] A. G. Worthing, *Am. J. Physics* **8**, 199 (1940).

$$1 \text{ ga-}\mu = 1 \text{ ab-}\mu = 1\frac{(\text{abpole})^2}{\text{dy cm}^2} = 1\frac{\text{abhen}}{\text{cm}}, \quad (24)$$

$$1 \text{ hl-}\mu = 4\pi\frac{(\text{hlpole})^2}{\text{dy cm}^2} = \frac{1}{4\pi c'^2}\frac{\text{hlhen}}{\text{cm}}, \quad (25)$$

$$1 \text{ MKS-}\mu = \frac{1}{4\pi}\frac{(\text{MKSpole})^2}{\text{nm}^2} = \frac{1}{4\pi}\frac{\text{hen}}{\text{m}}, \quad (26)$$

$$\mu_0 = \frac{1}{c'^2}\text{st-}\mu = 1 \text{ ab-}\mu = 1 \text{ hl-}\mu = \frac{4\pi}{10^7}\text{MKS-}\mu. \quad (27)$$

It is interesting, incidentally, that ϵ_0 and μ_0, using mks units (see note at end of paper), may be expressed by

$$\epsilon_0 = \frac{1}{9\times 10^9}\frac{\text{amp sec}}{\text{v m}} \quad (28)$$

and

$$\mu_0 = \frac{1}{10^7}\frac{\text{v sec}}{\text{amp m}}. \quad (29)$$

For the field of electromagnetism, a third basic relation is needed. Any one of several related equations, such as Ampere's, namely,

$$\mathbf{dH} = k'\frac{I\,\mathbf{dl}\times\mathbf{r}}{r^3}, \quad (30)$$

may be used. We prefer, however, to consider a relation derived therefrom and used by Jehle[4] in a related discussion, namely the magnetomotive force-electric current-linkage relation of Eqs. (1) to (5). Consider the equation

$$\oint \mathbf{H}\cdot\mathbf{ds} = k''nI. \quad (31)$$

Rearrangement yields

$$k'' = \oint \mathbf{H}\cdot\mathbf{ds}/nI. \quad (32)$$

For an evaluation of k'' in electromagnetic units, consider the case of a current I in a long straight wire which yields a magnetic field $\mathbf{H}$ of 2 oer at a distance of just 1 cm from the wire. From considerations closely connected with the definition of the abampere, the current is found to be just 1 abamp. Substitution of these values in Eq. (32) leads to the desired evaluation:

$$k'' = 4\pi \text{ oer cm/abamp turn}. \quad (33)$$

As for the proportionality constant k of Eqs. (8) to (12), we obtain with the aid of appropriate conversion factors the various convenient forms of expression for k'', namely,

$$k'' = 4\pi\frac{\text{oer cm}}{\text{abamp turn}}\frac{c' \text{ stoer}}{1 \text{ oer}}\frac{1 \text{ abamp}}{c' \text{ stamp}} = 4\pi\frac{\text{stoer cm}}{\text{stamp turn}}, \quad (34)$$

$$k'' = 4\pi\frac{\text{oer cm}}{\text{abamp turn}}\frac{1 \text{ abamp}}{c' \text{ stamp}} = \frac{4\pi}{c'}\frac{\text{oer cm}}{\text{stamp turn}}, \quad (35)$$

$$k'' = 4\pi\frac{\text{oer cm}}{\text{abamp turn}}\frac{1 \text{ hloer}}{\sqrt{4\pi} \text{ oer}}\frac{1 \text{ abamp}}{\sqrt{4\pi}\, c' \text{ hlamp}} = \frac{1}{c'}\frac{\text{hloer cm}}{\text{hlamp turn}}, \quad (36)$$

$$k'' = 4\pi\frac{\text{oer cm}}{\text{abamp turn}}\frac{(10^3/4\pi)(\text{amp turn/m})}{1 \text{ oer}}\frac{1 \text{ m}}{10^2 \text{ cm}}\frac{1 \text{ abamp}}{10 \text{ amp}} = 1, \quad (37)$$

$$k'' = 4\pi\frac{\text{oer cm}}{\text{abamp turn}}\frac{1 \text{ abamp}}{10 \text{ amp}} = \frac{4\pi}{10}\frac{\text{oer cm}}{\text{amp turn}}. \quad (38)$$

Collecting the various values for k'', we have

$$k'' = 4\pi\frac{\text{oer cm}}{\text{abamp turn}} = 4\pi\frac{\text{stoer cm}}{\text{stamp turn}} = \frac{4\pi}{c'}\frac{\text{oer cm}}{\text{stamp turn}} = \frac{1}{c'}\frac{\text{hloer cm}}{\text{hlamp turn}} = 1 = \frac{4\pi}{10}\frac{\text{oer cm}}{\text{amp turn}}. \quad (39)$$

[4] Jehle, *Am. J. Physics* **13**, 56 (1945).

The successive values collected in Eq. (39) represent the convenient forms for k'' when computations are based on the electromagnetic, the common electrostatic, the Gaussian, the Heaviside-Lorentz, the MKS and a common hybrid systems. Interestingly, the convenient form in the MKS system is the numeric 1. As above for k, it is obvious that k'' is a conversion factor with a value, as shown, equal to unity. It follows that, if the symbols represent physical quantities, the equation that suffices for the phenomenon is

$$\oint \mathbf{H}\cdot d\mathbf{s} = nI. \tag{40}$$

Comparison of the k'' values of Eq. (39) with the numerics 4π, $4\pi/c'$, $1/c'$, 1 and $4\pi/10$ of Eqs. (1) to (5) show that, if appropriate units of measurement are attached to those numerics, those equations will satisfy completely the "physical view." Answering the question asked earlier, "Are Eqs. (1) to (5) possibly on a par with Eq. (0) in omitting portions of conversion factors?" we say that they are. Furthermore, since all of the proportionality constants shown in Eq. (39) are equal to the numeric 1, Eq. (40) in its simple form with symbols representing physical quantities suffices to represent the magnetomotive force-electric current-linkage relation whatever the system of units one may use in making computations.

It cannot be said that the value of the proportionality constant will always be the numeric 1, though it always will be if the constant for the equation from any one "numeric view" is the numeric 1.

Consider next how Eq. (40) may be used to solve the simple problem of finding the current in a straight wire that will produce a magnetic field strength of 5 oer at a radial distance of 2 cm from the wire. Rearrangement, substitution of known values and application of the final conversion factor from Eq. (39) yields

$$I = \frac{\oint \mathbf{H}\cdot d\mathbf{s}}{n} = \frac{5\ \text{oer}\times 2\pi\times 2\ \text{cm}}{1\ \text{turn}}\,\frac{10\ \text{amp turn}}{4\pi\ \text{oer cm}} = 50\ \text{amp}, \tag{41}$$

in which the 10 amp turn/4π oer cm enters merely as a conversion factor.

Should one start with Eq. (29), the sole difference would consist of a k'' inserted with n in the second member, for k'' is precisely the conversion factor 4π oer cm/10 amp turn and is equal to unity. For other units used in expressing $\mathbf{H}$, $\mathbf{s}$ and I, other conversion factors from Eq. (39) would be used.

There are certain definite advantages for the practice of having the symbols of equations represent physical quantities. One is that the equations then take their simplest forms. Another is that a single equation, whatever the system of units that may be used for computations, then suffices to represent a phenomenon if the phenomenon is thus representable. The third and probably the main advantage is that the user may then think logically in terms of physical quantities rather than in terms of numerics. Whether or not these "physical view" advantages will suffice to overcome the present very general tendency to use the "numeric view" in the field of electricity and magnetism may be questioned. However, if adopted, it would seem certain that the "physical view" procedure would lead to clearer concepts of unity in nature than would result otherwise.

The writer gratefully acknowledges the helpful criticisms of his colleagues, O. H. Blackwood, F. A. Schwertz and W. C. Kelly and of a former colleague, J. J. Weigle, now of the University of Geneva.

Note added in proof.—Values for ϵ_0 and μ_0 as reported by users of the MKS system are respectively $1/4\pi$ times the value given in Eq. (28) and 4π times the value given in Eq. (29). Their values are to be looked on as measures of free space characteristics which are related to but different from the ϵ_0 and μ_0 characteristics of the other systems. The similarity and difference are much like those that would occur should one specify the size of a circle by its circumference rather than by its radius.

I know of no more encouraging fact than the unquestionable ability of man to elevate his life by a conscious endeavor.—THOREAU.

Forces on Ferromagnets through which Electrons Are Moving

David L. Webster
Stanford University, Stanford University, California

FORCES on permanent magnets through which conduction electrons are moving have been demonstrated qualitatively on lecture tables for a century or more. Usually the magnet is a cylindrical bar, and for the demonstration it is pivoted on pointed supports stuck in shallow holes in its ends, so that it is free to spin about its longitudinal axis. The current is introduced through one of these pivots and taken out by one or more brushes near the middle of the bar, or *vice versa*.

For many decades it was customary to explain the rotation of the bar in terms of forces caused by the magnetic field of the current, acting on all parts of the pole, considered as magnetic charge, through which the current ran. This explanation was always at variance with Ampere's theory of magnetism, and became definitely untenable when the truth of Ampere's theory was demonstrated by Barnett.[1]

The explanation which replaced it, being consistent with Ampere's theory, is attributed to Zeleny and Page.[2] This is based on Lorentz's principle that the average field strength within the iron is the flux density **B**, rather than the quantity **H**, which would have been that average if magnetic charges had existed. Conduction electrons, traveling in directions different from that of **B**, would naturally be deflected in each free path so that their impacts on the atoms would furnish the required forces.

This explanation contains a tacit assumption, that the local fields met by a conduction electron average up to the value **B**, that is, to the same value one would find by measuring field strengths at evenly spaced points throughout the iron, averaging them and then letting the distance between neighboring points approach zero, keeping them evenly spaced, and taking the limit of the average. A conduction electron, therefore, is treated in this explanation as having no special preferences between different parts of the atoms or regions between atoms, and as not in any way perturbing the Amperian currents.

Now, however, such a treatment of the conduction electron seems wrong. Wannier,[3] considering fast particles—electrons, mesons and so forth—shot through magnetized iron, presumably to measure their speeds, has shown that they perturb the Amperian currents far too much to let the fields they meet average to **B**. As proved by Barnett,[1] the Amperian currents are mostly electron spins. The strongest local fields in the direction of **B** are therefore within the magnetization electrons. So, if these electrons are pushed out of the way by invading electrons, the invaders must fail to meet these strongest local fields, and therefore must report (by their deflections) averages weaker than **B**. Protons, on the contrary, must report averages stronger than **B**. In either case, unless the invaders are going too fast to give the magnetization electrons time to get out of the way or into it, that is, unless they are going much faster than conduction electrons, the departures from **B** are large.

Evidently, now, there is still something seriously wrong with the explanation of the lecture experiment on the bar magnet, and something to be learned in getting it right. Moreover, it is not the experiment itself that is wrong. Zeleny and Page[2] settled this point. They not only calculated the torque on Lorentz's principle, but also measured it accurately and proved that the calculation gave the right value. In addition, they calculated the torque on the part of the circuit outside the magnet and, especially, on a very significant type of circuit in which the current was led into the bar magnet symmetrically, through a metal "apron" fastened around the middle of the bar. The torque on such an apron is opposite in direction to the torque on the magnet. If there are no sources of magnetic field but the bar and the current, and the radius of the apron is increased indefinitely, the calculated

[1] S. J. Barnett, *Science* **30**, 413 (1909); *Physical Rev.* **6**, 239 (1915); etc.; summarized in *Proc. Am. Acad. Arts Sci.* **75**, 109 (1944).

[2] J. Zeleny and L. Page, *Physical Rev.* **24**, 544 (1924).

[3] G. H. Wannier, *Physical Rev.* **67**, 364 (1945).

torque on it approaches the same magnitude as the calculated torque on the bar, as indeed it should, in the absence of radiation or transient effects, to conserve angular momentum. If the apron is nonmagnetic, there is no question of choice between **B** or **H** for it, so Zeleny and Page's calculation of its torque is indisputable. So also, therefore, is their value for the torque on the magnet. It is, in fact, the torque that would be produced by a force per unit volume equal to $\mathbf{i}\times\mathbf{B}$, where **i** is the current density, just as though the average field strength encountered by every conduction electron was **B** and there were no other sources of torque.

If conduction electrons do not experience as much force as this, some other forces must make up for the deficiency. The next questions, therefore, are: (i) how much deficiency is there; and (ii) what forces make up for it?

Wannier's Equation

According to Wannier,[3] if an invading "test charge" is much more massive than an electron (presumably so that it will move practically straight and steadily) the average strength **b** of all the fields it meets in a ferromagnet (with its magnetization **M** due almost entirely to electron spins) is

$$\mathbf{b}=\mathbf{H}+2\pi(1+p)\mathbf{M}=\mathbf{B}-2\pi(1-p)\mathbf{M}, \quad (1)$$

where

$$p=x/(1-e^{-x}), \quad (2)$$

$$x=4\pi^2Ze^2/hv. \quad (3)$$

Here v is the velocity of the charge, Ze is its quantity ($-e$ for an electron, $+e$ for a proton, and so forth), and p is the relative probability of coincidence of the test charge and a magnetization electron, as compared to randomness.

For velocities as low as those of conduction electrons, $p<0.01$; so, practically,

$$\mathbf{b}=\tfrac{1}{2}(\mathbf{B}+\mathbf{H}). \quad (4)$$

Conduction electrons, to be sure, do not satisfy the afore-mentioned mass requirement. Therefore, along with pushing magnetization electrons out of their way, to some extent, they themselves must be deflected so as to tend to pick their way between the magnetization electrons. For them, p must be even smaller than the value given by Eqs. (2) and (3); so Eq. (4) must hold even better than if they were massive.

This equation is indeed the same as one found by Swann[4] for cosmic-ray particles going through not too thick ferromagnets, on the assumption that the magnetization electrons could be treated as strictly classical spinning electrons, with the classical electron radius. On this basis, even in a magnet a whole centimeter thick most cosmic-ray particles would miss all the magnetization electrons, just because of their extreme smallness. From a wave-mechanical viewpoint such dimensions must not be taken so literally. According to Wannier,[3] indeed, a "head-on collision" could be expected about once in every hundred atoms penetrated. From this viewpoint it is electrostatic repulsion, rather than small dimensions, that is the basis for Eq. (4).

Regardless of the basis, however, the result is that when a current is conducted through a ferromagnet for which $\mathbf{H}\ll\mathbf{B}$, the force per unit volume on the conduction electrons is only about half of $\mathbf{i}\times\mathbf{B}$, which was the value apparently so well confirmed by Zeleny and Page's measurements. On what, then, does the other half of the measured force act?

Comparison with Conductors in Motors

The obvious guess on this question is the correct one: the rest of the force acts on the magnetization electrons which were pushed out of the way. Indeed, the situation of each conduction electron is closely analogous to that of electrons in the conductors on a slotted armature of a motor. The torque on such an armature can be calculated correctly by assuming the field strength at the wires to be the average of **B** for the whole air gap, just as if there were no slots and teeth, then calculating the forces this field would exert on the wires and neglecting the forces on the teeth. To prove this algebraically it is easiest to reason in terms of induced emf's and power; but to see how it happens one has only to notice that the weakening of the field in the slots and the strengthening in the teeth are caused by

[4] W. F. G. Swann, *Physical Rev.* **49**, 574 (1936).

uncancelled Amperian currents in the sides of the teeth such as are represented in Fig. 1.[5] These currents run parallel and antiparallel to the currents in the wires, so the mutual attractions and repulsions simply transfer a large part of the force from the wires to the teeth, without any appreciable change in the resultant force associated with each wire.

Returning to conduction electrons in a ferromagnet, if such an electron succeeds entirely in either pushing the magnetization electrons out of its way or picking its way between them, so that it is never coincident with any of them—that is, if its relative coincidence probability p is zero—then the situation is practically the same as if a fine hole had been drilled through the magnet and a wire run through it to conduct this electron. Therefore the rule that governs the transfer of a part of the force from a wire in a motor to the adjacent teeth should dictate a similar transfer from this electron to the adjacent magnetization electrons. The total force on the magnet, associated with this conduction electron, should therefore be approximately the same as if it were alone

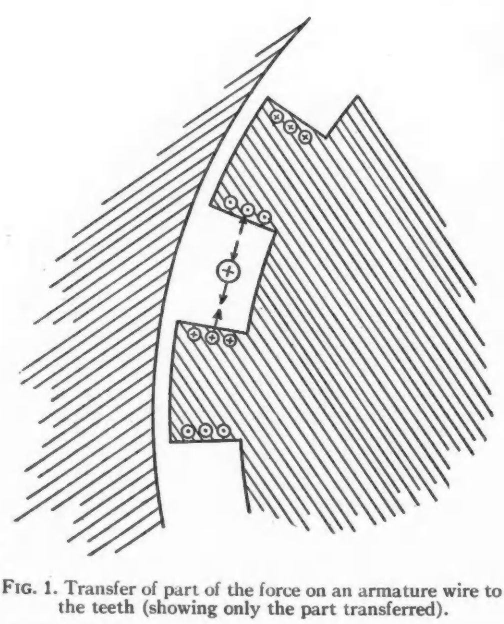


FIG. 1. Transfer of part of the force on an armature wire to the teeth (showing only the part transferred).

[5] From D. L. Webster, *Am. J. Physics* (*Am. Phys. T.*) **2**, 7 (1934).

in a uniform field of strength **B**. For conduction electrons with p not quite reduced to zero, reasoning along this line may not be quite so clear; but it is clear enough to indicate that the force per unit volume probably should be at least roughly equal to $\mathbf{i}\times\mathbf{B}$.

Conversely, since Zeleny and Page[2] found that the force per unit volume did indeed equal $\mathbf{i}\times\mathbf{B}$ as accurately as it could be measured, and since Wannier's equation gives the conduction electrons only about half this measured force, the foregoing reasoning points to the neighboring magnetization electrons as the recipients of its other half. In short, the explanation of forces on magnets through which currents flow is not quite so simple as we had thought; but it is no worse than the explanation of the torque of a motor. In fact, the two explanations are essentially the same.

Ferromagnets with Fast Test Charges

If conduction electrons give rise to these forces on the magnetization electrons, what will fast test charges do when they traverse a magnet? To answer this question it will be well to look first at a proof of the law for conduction electrons, essentially equivalent to that of Zeleny and Page,[2] but more readily modified to apply to fast test charges.

Figure 2 represents a vertical steel plate, of height h, width w and thickness t, with uniform, vertically upward magnetization $\mathbf{M}$; h and w may have any arbitrary ratio, but t must be very small in comparison with h. Touching the middle of each side of the plate, and perpendicular to it, is an infinitely long straight wire that carries a current I, directed away from us.

For purposes of mathematical rigor these wires may first be given a finite length l, and the circuit may be closed with a semicircular piece of wire. Then it can be shown that if l becomes infinite, the force on the piece of wire approaches zero, and so does its contribution to the magnetic field at the plate. It is allowable, therefore, to let l be infinite and consider I to be constant, both in time and in space along the wires. Then Newton's third law of motion can be applied, with the wires considered as one body and the plate as the other.

The force on the plate may be divided into two parts: F_1, the force on the conduction electrons and neighboring magnetization electrons where the current goes through the plate; and F_2, the force on all uncancelled Amperian currents in the surfaces of the plate. The force on the wires will be called F_3. All forces will be considered positive if acting to the right. With these definitions, Newton's law has taken the form

$$F_1+F_2=-F_3. \tag{5}$$

If t is negligible in comparison with h, both F_2 and F_3 are easily calculated, with the results that

$$F_2=-8ItM\tan^{-1}(h/w) \tag{6}$$

and

$$F_3=-8ItM\tan^{-1}(w/h); \tag{7}$$

therefore

$$F_1=+4\pi ItM=ItB. \tag{8}$$

This proof may now be modified so as to apply to fast test charges. In place of the contacts of the wires to the plate, let us assume that the farther wire terminates in an electron gun whose transverse dimensions are all negligible in comparison with either h or w, and that the nearer wire terminates in a Faraday cylinder which catches the electrons coming through the plate. (Secondary emission and captures of electrons are neglected here.) Further, let us assume the gun to be so directed that the transverse component of momentum of each electron shot from it is simply reversed in direction by the field in the plate, without change of magnitude; and let the length of the electron beam be too short for appreciable deflection by the field outside the plate.

Considering only transverse components of forces, F_2 and F_3 are just as before, so Eq. (8) still gives the total of the forces on the gun electrons and neighboring magnetization electrons added together. But now it is no longer right to add them. The magnetization electrons obviously belong to the plate, and any force F_m acting on them is a part of the force we could measure by holding the plate with some sort of balance. The gun electrons do not belong to the plate in any such way. Neither do they belong to it in the sense of conduction electrons, because they do not transfer to it the transverse momentum they acquire while in it. Instead, they

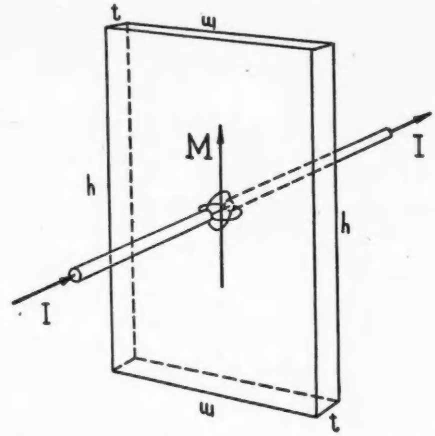


FIG. 2. Magnetized plate traversed by a current.

transfer this momentum to the wires, through recoil as they leave the gun and impact as they arrive in the Faraday cylinder. The force on the gun electrons therefore belongs to the wires. Calling it F_g, we have Newton's law now taking the form

$$F_m+F_2=-F_g-F_3. \tag{9}$$

By Wannier's equation, Eq. (1),

$$F_g=Itb=It[B-2\pi(1-p)M]. \tag{10}$$

Combining the last five equations, we obtain

$$F_m=2\pi(1-p)ItM. \tag{11}$$

Since H is in this set-up small in comparison with B, Eqs. (10) and (11) may be written approximately as

$$F_g=\tfrac{1}{2}(1+p)ItB \tag{12}$$

and

$$F_m=\tfrac{1}{2}(1-p)ItB. \tag{13}$$

Since $p<1$ for electrons, both of these forces are in the same direction. For positrons and protons, on the contrary, $p>1$, so that with them used as test charges F_g and F_m would have opposite directions.

A good question now is whether F_m can actually be found by experiments with an electron gun and a plate. If so, the best set-up naturally would include a gun of fairly high velocity, so that there will not be too much heating of the plate, and for the same reason the plate should be a very thin

foil. Two such foils, oppositely magnetized, might be attached to opposite ends of a light metal frame hanging on a Wollaston wire, which will carry away any electric charges. Then if an electron beam is shot through both foils, the frame ought to turn.

It must be remembered, however, that the resultant force on each foil is not F_m, but F_m+F_2 as in Eq. (9). This sum is

$$F_m+F_2=ItM[2\pi(1-p)-8\tan^{-1}(h/w)]; \quad (14)$$

so it may not even have the direction of F_m, and indeed will not unless

$$(h/w)<\tan[\tfrac{1}{4}\pi(1-p)]. \quad (15)$$

Practically there may be considerable difficulty in making an experiment of this sort. The shape suggested by Eq. (15) is unfavorable, since it requires a relatively wide frame and this introduces undesirable weight, requiring a fairly strong Wollaston wire to support it. Preliminary calculations indicate that one might get a readable deflection with about 1 ma of electrons going through foils 1 micron thick, provided the heat produced by the electrons did not demagnetize the foils. The probable error would be large, however, so no promises are made here about any such experiment.

Rather, the definite conclusions here are: (i) that with fast electrons as well as with conduction electrons, the force F_m on the neighboring magnetization electrons must exist; and (ii) that with fast electrons F_m acts on the magnet while the force F_g on the fast electrons does not, so that it should be possible, in principle and perhaps also in practice—though not on the lecture table!—to separate them and demonstrate the existence of F_m by direct mechanical measurement.

A Definition of Temperature as a Secondary Quantity

AUSTIN J. O'LEARY
The City College, New York 10, New York

IT is my purpose to indicate how the physical quantity *temperature* may be defined in statistical theory without reference to a thermal standard of any kind and how all thermodynamic measurements may be based directly on statistical theory. I shall then discuss the definition of temperature from the macroscopic point of view.

Since this subject is highly controversial, it seems advisable to state the meaning of a few terms as they are here used.

(1) *Standards.*—A standard is *nonessential* if it is introduced for practical convenience as a supplement to the primary standards. The following are examples: standard candle, standard cell, standard resistor, water and mercury. Water is used as a standard substance in defining specific gravity, specific heat and the practical heat units; mercury is used as a standard substance in defining certain practical units of pressure and of electric resistance. Examples of thermal standards will be given in the last section. I shall indicate that thermal standards are nonessential. If this view is accepted for the time being at least, we are left with four types of primary standards: two kinematic, one dynamic and one electrodynamic. The particular standards in most general use are the International Meter, the earth in its motion with respect to the sun and the stars, the International Kilogram and space (a standard medium).

(2) *Physical quantities.*—A physical quantity is any explicitly defined measure used in describing physical phenomena. Four *primary quantities* and corresponding *standard magnitudes* are defined with reference to four different types of measurement involving the primary standards. Units are defined by assigning specific values to the standard magnitudes.

The four standard magnitudes and the values assigned to them are: (i) the rest length between two parallel markings on the International Meter under standard pressure at 0°C ($l_0=1$ m$=10^2$ cm$=10^{-3}$ km, etc.); (ii) the interval of time between two successive transits of the "fictitious sun" across a meridian ($t_0=1$ mean solar day $=86{,}400$ sec, etc.); (iii) the rest mass of the International Kilogram ($m_0=1$ kgm$=10^3$ gm$=10^6$ mgm, etc.); (iv) the

permeability of space ($\mu_0 = 1$ cgs-em unit $= 1/c^2$ cgs-es unit $= 10^{-7}$ mks unit, where 1 mks unit $= 1$ henry $m^{-1} = 1$ kgm m $coul^{-2} = 1$ newton amp^{-2}, etc.).

The choice of primary quantities is dictated largely by convenience; an example is the choice between permeability and permittivity; permeability happens to be the quantity most directly involved in measuring the absolute values of practical standards.[1] All other quantities that are needed can be defined as secondary quantities in terms of the four primary quantities. In most cases, a secondary quantity is nothing more than a particular combination of primary quantities that is given a special name; an equation expressing the definition is a simple identity. A secondary quantity may also be defined by an equation that is not a simple identity; the equation may have any form as long as the quantity to be defined is uniquely determined in terms of quantities already defined; to become an accepted *physical quantity*, a quantity so defined must prove to be a useful physical measure. Similarly, two quantities may be defined by two equations involving both quantities, or one quantity may be defined and a physical relation stated in two equations that do not have meaning independently of one another, and so forth.

The units of a secondary quantity are automatically expressed in terms of whatever units are used for the primary quantities; they are independent of any particular method of measuring the quantity and are independent of any standard other than the four primary magnitudes or a magnitude fixed in terms of these, such as the standard magnitude of acceleration, g_s. That is what is meant by a *secondary quantity*. For the sake of convenience, we sometimes define additional practical units for a secondary quantity appropriate to a particular method of measurement.

Statistical Method

The interpretation of statistical equations best suited to my purpose is that outlined recently by Schrödinger.[2] Some indication of his view is given by the following quotation:

> There is, essentially, only one problem in statistical thermodynamics: the distribution of a given amount of energy E over N identical systems. Or perhaps better: to determine the distribution of an assembly of N identical systems over the possible states in which this assembly can find itself, given that the energy of the assembly is a constant E. The idea is that there is weak interaction between them, so weak that the energy of interaction can be disregarded, that one can speak of the "private" energy of every one of them and that the sum of their "private" energies has to equal E. . . . Here the N identical systems are mental copies of the one system under consideration—of the one macroscopic device that is actually erected on our laboratory table. Now what on earth could it mean, physically, to distribute a given amount of energy E over these N mental copies? The idea is, in my view, that you can, of course, imagine that you really had N copies of your system, that they really were in "weak interaction" with each other, but isolated from the rest of the world. Fixing your attention on one of them, you find it in a peculiar kind of "heat bath" which consists of the $N-1$ others.

Let $a_1, a_2, a_3, \cdots, a_l, \cdots$ be the numbers of systems in the energy states $\epsilon_1, \epsilon_2, \epsilon_3, \cdots, \epsilon_l, \cdots$, where each system is distinguishable from each other system; then the number of single states belonging to this class of the assembly is

$$P = \frac{N!}{a_1!a_2!a_3!\cdots a_l!\cdots}, \tag{1}$$

where $\sum_l a_l = N$, and $\sum_l a_l\epsilon_l = E$. By two different methods, each based upon a slightly different probability postulate, we can get the basic distribution

$$a_l = \frac{N \exp(-\epsilon_l/\theta)}{\Sigma \exp(-\epsilon_l/\theta)} \equiv -N\theta \frac{\partial}{\partial \epsilon_l} \log \Sigma \exp(-\epsilon_l/\theta), \tag{2}$$

where θ, the same for all N systems, is an intensive scalar quantity or parameter that distinguishes one assembly of equilibrium states from another; it is a positive parameter with limiting value zero, units having the dimensions $[ML^2T^{-2}]$, and is related to the average share of energy of one system, $U = E/N$, by the equation

$$U = \frac{\Sigma \epsilon_l \exp(-\epsilon_l/\theta)}{\Sigma \exp(-\epsilon_l/\theta)} \equiv \theta^2 \frac{\partial}{\partial \theta} \log \Sigma \exp(-\epsilon_l/\theta). \tag{3}$$

These equations might equally well be expressed in terms of any function of θ. In reference 2, the reciprocal quantity μ is used because the method

[1] L. J. Briggs, "The national standards of measurement," *Rev. Mod. Physics* 11, 111–120 (1939).

[2] E. Schrödinger, *Statistical thermodynamics* (Cambridge Univ. Press, Macmillan, 1946).

of most probable distribution happens to be presented first (Chapter II), in which method μ enters as a Lagrange multiplier. In the Darwin and Fowler method of mean values (Chapter VI, reference 2), a parameter $z[\equiv e^{-1/\theta}]$ enters as the saddle point in the mathematical method of steepest descent. The quantity θ has always been used in the Gibbs ensemble, and I am using it in this paper because its values are so much more convenient than those of μ or z.

The defining Eqs. (2) and (3) naturally depend upon the nature of the thermodynamic system of interest. They can be solved explicitly for particular kinds of systems, but no general solution can be obtained for all kinds of systems. However, independently of any particular solution, one can deduce, for all systems, exact analogs of the basic laws of thermodynamics in all their varied forms.[3] From the correspondence between these two types of relations, it has been traditional procedure to identify θ as being equivalent to the thermodynamic quantity kT and some such quantity as

$$\log \Sigma P/N[=\log P_{\max}/N]$$

as being equivalent to S/k, where S is thermodynamic entropy. I wish to suggest a more direct procedure.

The need for a thermal standard is eliminated by the addition of a statistical postulate to mechanics, thereby extending the scope of mechanics to the field of thermodynamics. Let us therefore reconstitute experimental thermodynamics in terms of quantities defined in statistical theory, whence the aforementioned statistical theorems with minor changes in phraseology become derived physical laws subject to *a posteriori* confirmation. Let us call θ the *statistical temperature* of a system on the *primary scale;* and let us define other scales of statistical temperature by the equation

$$\Theta=\theta/k_\theta, \tag{4}$$

where k_θ is any convenient numeric fixed by convention; the primary scale is a special case of Eq. (4) in which $k_\theta=1$. Let us define an extensive dimensionless scalar quantity S_θ, which we may call the *statistical entropy* of a system, by the equation

$$S_\theta=k_\theta \log \Sigma P/N. \tag{5}$$

In terms of quantities such as these, defined for actual thermodynamic systems, statistical theory gives a full statement of the basic laws of thermodynamics in a variety of forms, including a description of processes such as the Carnot cycle.

We now face the important task of relating theory to experiment, step by step, with no preconceived judgments. The first step is the establishment of methods of measuring θ. Several derived laws afford ready means of measurement.

(1) *The Planck radiation law and its corollaries.*—The radiation laws contain only two constants, h and c, each of which is known independently of these particular laws. Hence, within the limits of experimental error, blackbody radiation curves provide a test of the Planck law, and each different curve yields its own unique value of θ. Each value of θ so measured must, of course, agree with that obtained by any other appropriate method, otherwise a flaw in theory is revealed. Independent measurements based upon different laws are in fact available as a check upon the self-consistency of the theory.

Planck, in his original paper, followed a reverse procedure; he arrived at a value of h from blackbody measurements at known temperatures, using the data reported a few months earlier by Lummer and Pringsheim.

(2) *The gas law.*—In the case of an ideal monatomic gas, we get

$$PV/M=N_0\theta=\Theta', \tag{6}$$

where P is the pressure, V the volume, M the number of moles, N_0 the Avogadro number, and Θ' the temperature on what I shall call the *Avogadro scale*, defined by setting $k_\theta=N_0^{-1}$; the prime is used to identify Θ with this particular scale. Equation (6) applies also to an ideal polyatomic gas for which there is equipartition of energy among the various degrees of freedom. It is to be expected from theory that any actual gas at sufficiently low pressure and density should behave like an ideal gas. This expectation can

[3] Details are presented in several textbooks; in particular, see R. C. Tolman, *The principles of statistical mechanics* (Oxford, 1938), Chap. XIII.

be verified by experiment. Also, the existence of a number of invariable temperatures can be demonstrated and their values measured for use as reference points. In these experiments, we apply the law of thermal equilibrium: different systems are in thermal equilibrium with one another if, and only if, θ is the same for each. For an enclosed gas in equilibrium with an ice bath under standard pressure, it can be shown that $\lim_{P\to 0} PV/M$ is the same for all gases, independently of the circumstances of time, place, humidity, and so forth; similarly for the steam point and various melting and boiling points, some of which can be checked by radiation measurements. Thus Θ' can be determined from measurements of pressure and molal volume of an actual gas by means of the relation

$$\Theta' = \lim_{P\to 0} PV/M. \qquad (7)$$

Let Θ_0' and Θ_1' denote the temperatures of the ice point and steam point, respectively. Experiment yields the following values:[4]

$$\begin{aligned}\Theta_0' &= 2\,271.15 \pm 0.06 \text{ kgm m}^2\text{ sec}^{-2}\\ &\equiv 1\,675.11 \pm 0.04 \text{ lb ft;}\\ \Theta_1' &= 3\,102.59 \pm 0.18 \text{ kgm m}^2\text{ sec}^{-2}\\ &\equiv 2\,288.35 \pm 0.13 \text{ lb ft.}\end{aligned}$$

Dividing by the known value of N_0, namely, $(6.0228 \pm 0.0011) \times 10^{23}$, we get

$$\begin{aligned}\theta_0 &= (3.7709 \pm 0.0007) \times 10^{-21} \text{ kgm m}^2\text{ sec}^{-2},\\ \theta_1 &= (5.1514 \pm 0.0031) \times 10^{-21} \text{ kgm m}^2\text{ sec}^{-2}.\end{aligned}$$

[4] R. T. Birge, *Rev. Mod. Physics* **13**, 233 (1941). For the methods of evaluating physical constants, see R. T. Birge, *Rev. Mod. Physics* 1, 1 (1929). The value of Θ_0' is obtained from pressure-volume measurements on oxygen at the ice point; T_0, equal to 273.16 ± 0.01°K, is derived from measurements of the pressure coefficient and volume coefficient of a gas, and from measurements involving the Joule-Thomson effect. I have computed the above value of Θ_1' from the relation $\Theta_1' = \Theta_0'(T_0+100)/T_0$; Θ_1' might, of course, be determined directly from pressure-volume measurements on a gas at the steam point.

Although the units of θ have the same dimensions as energy units, $[ML^2T^{-2}]$, θ differs from energy in this one respect: it is an intensive rather than an extensive quantity. There is a comparable difference between the vector quantity torque and the scalar quantity energy; their units have the same dimensions, yet they represent different physical quantities. Torque units are usually written in a slightly different way from those of work and energy; the name "joule" is not applied to the mks unit of torque, nor the name "erg" to the cgs unit. In similar fashion, it may be well to maintain some slight distinction between energy and the parameter θ by writing the units of θ in the form "kgm m² sec⁻²," "gm cm² sec⁻²," and "lb ft."

TABLE I. Characteristics of five absolute temperature scales of the type defined by Eqs. (9) and (10). With the exception of two integers, 100 and 180, the numerical values in the table represent the following approximations: $8.31436 \pm 0.00038 \simeq 8.3$; $3.4068 \pm 0.0002 \simeq 3.4$; $273.16 \pm 0.01 \simeq 273$; $491.69 \pm 0.02 \simeq 492$.

	Scale	C	R	Ice point
Θ	mks	100 kgm m² sec⁻²	8.3	273 kgm m² sec⁻²
	cgs	100 gm cm² sec⁻²	8.3×10^7	273 gm cm² sec⁻²
	B.E.	180 lb ft	3.4	492 lb ft
T	Kelvin	100	8.3 kgm m² sec⁻² (°K)⁻¹	273°K
	Rankine	180	3.4 lb ft (°*R*)⁻¹	492°*R*

From the law of thermal equilibrium, it follows that the gas thermometer can be used to measure the statistical temperature of any system in equilibrium with it and to calibrate other types of thermometers.

As the next step in relating theory to experiment, we may define practical units of heat quantity and measure the Joule equivalent. The whole field of thermodynamic measurements, with a foundation in statistical theory, is then opened up. The remaining steps are apparent. All the derived laws can be confirmed by experiment, and this verification confirms the statistical postulate from which the laws are deduced.

It is convenient to have statistical temperature scales in which $\Theta_1 - \Theta_0 = C_\theta$, where the numerical value of C_θ is a chosen integer. Scales of this special type are defined by setting

$$k_\theta = (\theta_1 - \theta_0)/C_\theta, \qquad (8)$$

or, since $\Theta'[\equiv N_0\theta]$ is measured more directly than θ, by setting

$$R_\theta = N_0 k_\theta = (\Theta_1' - \Theta_0')/C_\theta. \qquad (9)$$

Each scale so defined has its own characteristic value of C_θ, k_θ and R_θ. Unlike the primary and Avogadro scales, each scale defined by Eq. (9) has a single characteristic unit. Three such absolute scales are indicated in the first part of Table I. For many purposes, relative scales are more convenient; a centigrade scale can readily be defined in terms of the mks or cgs scale, and a Fahrenheit scale in terms of the British Engineering scale.

In place of θ, one might equally well develop experimental thermodynamics in terms of μ or some other function of θ. A system that corresponds closely to the usual form of thermo-

dynamics is formulated as follows: constants R_T and k_T for the purpose of specifying particular scales are defined by the equation

$$R_T = N_0 k_T = (\Theta_1' - \Theta_0')/C_T, \tag{10}$$

where C_T is an integer, equal to 100 for the Kelvin scale and 180 for the Rankine scale. *Thermodynamic temperature* T, such that $T_1 - T_0 = C_T$, and *thermodynamic entropy* S_T are defined by the equations

$$T = \theta/k_T, \tag{11}$$

$$S_T = k_T \log \Sigma P/N. \tag{12}$$

Since C_T is arbitrarily taken to be an integer in these definitions, the units of k_T, R_T and S_T are thereby assigned the dimensions $[ML^2T^{-2}]$ and the unit of T is made dimensionless. The *degree Kelvin* and *degree Rankine* are then dimensionless units similar to the degree in angular measure. The two thermodynamic temperature scales are listed in Table I for easy comparison with the three similar scales of statistical temperature. Combining Eqs. (7), (9) and (10), we get the following expression for the constants R_T and R_θ in Table I:

$$R_T(\text{or } R_\theta) = \frac{\lim_{P\to 0}(PV/M)_1 - \lim_{P\to 0}(PV/M)_0}{C_T(\text{or } C_\theta)}. \tag{13}$$

The gas law, Eq. (7), may be rewritten in the following form appropriate to all statistical and thermodynamic scales:

$$\lim_{P\to 0}(PV/M) = R_T T = R_\theta \Theta. \tag{14}$$

The various constants R assume significance with respect to a mole of gas, and the corresponding constants $k[\equiv R/N_0]$ with respect to a gas molecule, only because of the way in which these constants occur in Eq. (14). All are defined without any reference to a gas: for the scales in Table I, R and k are defined by Eqs. (9) and (10); for the primary scale, $k_\theta = 1$, $R_\theta = N_0$; and for the Avogadro scale, $k_\theta = N_0^{-1}$, $R_\theta = 1$. The two quantities T and Θ serve the same purpose; they differ only as the result of a different arbitrary assignment of units and dimensions in the two cases; similarly for S_T and S_θ, R_T and R_θ, k_T and k_θ.

There would seem to be several advantages to be gained from adoption of the foregoing procedure:

(i) The laws of thermodynamics become deducible from theory; in other words, the relations deduced in statistical mechanics become actual physical laws, not mere analogues of laws. Instead of being merely a supporting theory, statistical mechanics, together with the experiments based upon it, becomes a full-fledged theory of thermodynamics.

(ii) The various constants k are defined without any reference to a gas. There remains no occasion for surprise at the appearance of k or kT in numerous relations wholly unrelated to a gas.

(iii) Temperature becomes, unequivocally, a secondary quantity. I think it should be so defined in view of the marked difference between thermodynamic properties, on the one hand, and the more fundamental dynamic and electrodynamic properties of matter, on the other hand. The latter are universal characteristics of matter, whereas the former are exhibited by macroscopic bodies only. Unlike quantities such as mass and electric charge, temperature and other quantities peculiar to thermodynamics have no meaning in the case of individual fundamental particles.

The Macroscopic Point of View

The successive steps in a general definition of temperature from the macroscopic point of view are as follows:[5] introduction of a generalized force and generalized displacement to specify thermodynamic states of a system; definition of what is meant by thermal equilibrium, and introduction of the zeroth law of thermodynamics (the law of thermal equilibrium) as an empirical law; establishment of sets of corresponding isotherms for different systems such that *a single state of thermal equilibrium* is common to all thermodynamic states of corresponding isotherms of all systems; and finally, the definition of a physical

[5] The procedure has been carefully described in detail by L. Balamuth, H. C. Wolfe and M. W. Zemansky, "The temperature concept from the macroscopic point of view," *Am. J. Physics* 9, 199 (1941). I should perhaps point out that my view concerning the final step differs somewhat from theirs. Since the word "temperature" is used as the name of a physical quantity, I prefer not to use the same word in a second sense, as the name of a property. Also, I regard the definition of units and dimensions to be an inseparable part of the definition of any physical quantity.

quantity to characterize *different states of thermal equilibrium*. The quantity $\lim_{P\to 0}(PV/M)$, by itself or multiplied by any constant, is a convenient measure of the state of thermal equilibrium of a gas and of any system in equilibrium with a gas. The gas constant R may be defined by Eq. (13) and thermodynamic temperature T by Eq. (14). In doing so, we introduce a thermal standard, namely, a gas under low pressure—a standard type of system. Fixed points are not used as standards in this definition; they are employed only to define convenient values of R. If parameters other than P and V were to be used, we would be limited still more narrowly in our choice of a standard system.

A thermal standard is also needed in a definition of temperature according to the Kelvin method. One may specify a given temperature difference between two fixed points, or assign a particular temperature T_s to one fixed point. In the latter case, it is ideally possible to measure any temperature T uniquely and independently of the nature, phase and mass of the thermometric system, in accordance with the defining equation

$$T = T_s Q/Q_s.$$

There seems to be no way of defining temperature from a purely macroscopic approach without employing a thermal standard of some kind; and if one were to maintain this approach without regard to the internal structure of matter and were to follow the same procedure as in dynamics and electrodynamics, one would treat temperature as a primary quantity and assign to its units a fifth physical dimension. However, the need for a thermal standard in the macroscopic approach appears to be another of the limitations in this point of view. We do not get far toward an understanding of thermal phenomena until we recognize the particle structure of matter, and then it is seen that thermal properties are characteristic of particles in aggregate, but not of individual particles, and we learn how to describe accurately the behavior of the aggregate in terms of statistical parameters.

If the statistical method presented here is accepted, thermal standards are nonessential and should be so recognized even when, as a matter of convenience, attention is restricted to a purely macroscopic description. Naturally, this conclusion has nothing to do with the use of thermal standards for practical purposes. It merely makes it desirable to apportion the dimensions $[ML^2T^{-2}]$ between the units of R and T without introducing a fifth dimension; it would seem best to make the units of T dimensionless and assign the dimensions $[ML^2T^{-2}]$ to the units of R, k and S.

The Determination of Driving Forces Required to Produce Specified Motions

D. A. Wells
University of Cincinnati, Cincinnati, Ohio

THE type of problem most frequently treated in textbooks on dynamics is one in which the motions of a system are to be found from given applied forces. The converse problem, in which the motions are prescribed and the driving forces necessary to produce such motions are to be found, is rarely considered except in connection with rather elementary examples.

The purpose of this paper is to present a general method of handling the latter type of dynamical problem. The generality of the method presented and the ease with which it can be applied reduces the task of finding driving forces to certain routine operations, even for very complex systems and assumed motions.

Preliminary Considerations

As a preliminary step toward the final development, careful consideration should be given to the following matters.

(a) *Possible independent motions of a system.*—Since any one of the generalized coordinates

$q_1, q_2, \cdots, q_n$ representing the configuration of a holonomic system having n degrees of freedom can be varied independently, it is convenient to refer to the system as having n possible "independent motions." That is to say, the most general type of motion of which a system is capable (consistent with constraints) can be regarded as made up of n independent motions along the n independently variable coordinates. Or again, from the point of view of prescribed motions, each of the coordinates $q_1, q_2, \cdots, q_n$ can be assumed to change with time in any manner whatever and the result will be a possible motion of the system. Moreover, since the coordinates of a system can be choosen in various ways, the independent motions of a system can be regarded as the motions associated with any one of various sets of coordinates. However, the number of motions that can be assigned arbitrarily remains always equal to the number of degrees of freedom of the system.

(b) *Ways in which motions can be expressed.*—From the statements of the preceding paragraph it is clear that any motion of a system can be represented by expressing each coordinate as a function of time. Implicitly such expressions take the form

$$q_1 = \varphi_1(t),\ q_2 = \varphi_2(t),\ \cdots,\ q_n = \varphi_n(t). \qquad (1)$$

To represent a specific motion of a system the functions must, of course, be properly chosen. For example, if a particle that is free to move in a plane (two degrees of freedom) is to be made to move with constant velocity along the straight line $y = bx$, then $y = v_1 t$, $x = v_2 t$ represent the desired motion for $v_1/v_2 = b$. Or again, relations

$$x = r\cos(vt/r), \quad y = r\sin(vt/r)$$

represent the motion of a particle in a circle of radius r with constant speed v.

(c) *Possible ways of applying driving forces.*—If n independently controlled forces are applied to a system of n degrees of freedom in such a manner that any one of the independently variable coordinates can be altered by any one or more of the forces, without necessarily altering any of the other coordinates, values of these forces can be found that will produce any prescribed motions. Driving forces can usually be applied to a system in a number of ways, any one of which will achieve the desired results. The fundamental idea of this paragraph can best be shown with a specific example.

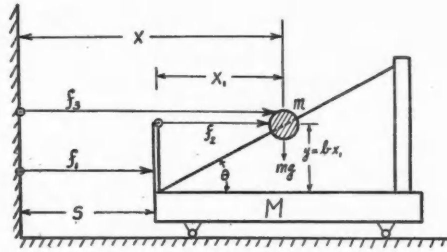


FIG. 1. System for illustrating basic principles involved in finding driving forces.

In Fig. 1, an object of mass m is free to move along a smooth inclined rod supported rigidly on a cart, the total mass of which is M. The cart is free to move along a straight horizontal track. By applying proper forces the cart can be given any desired horizontal motion at the same time that m is made to move in any manner along the rod. There are two degrees of freedom and hence two independently controlled forces are required. Among the possible sets of coordinates, any one of which can be used to represent the configuration of the system, are (x_1, s), (x_1, x) and (x, s). A little consideration will show that any one of the sets of forces (f_1, f_2), (f_1, f_3), (f_2, f_3) will produce any desired motions, provided, of course, that each force in the chosen set is given the proper value. Note that the forces of any one of the sets fulfill the requirements stated in the first sentence of this section. If one at first has doubts as to whether f_2 and f_3 alone can be found so as to give M and m any desired motions, these will be dispelled by examples to follow.

(d) *Location of driving mechanisms.*—Since for every "action" there is an equal and opposite "reaction," it is important to specify where the mechanism producing a particular driving force is located. In the figures a circle attached to the tail of a force arrow indicates the driving mechanism. For example f_2, Fig. 1, is produced by a device attached to the cart, while f_1 and f_3 are produced by driving mechanisms attached to the frame of reference.

Outline of General Theory

The theory here presented is based on a special use of the Lagrangian equations[1] of motion, the derivation of which will not be given. For a system having n degrees of freedom these equations can be written as

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_r}\right)-\frac{\partial T}{\partial q_r}=Q_r, \tag{2}$$

where T, the kinetic energy of the system, is expressed in terms of the $q_1, q_2, \cdots, q_n$ independently variable coordinates and their time derivatives $\dot{q}_1, \dot{q}_2, \cdots, \dot{q}_n$. Time may enter T explicitly if there are moving coordinates and (or) moving constraints. The quantity Q_r is referred to as a *generalized* force. Specific expressions for the generalized forces are obtained from the relation

$$\delta W_{q_r}=Q_r\delta q_r, \tag{3}$$

where δW_{q_r} is the work done by any and all forces when the coordinate q_r is given a displacement $+\delta q_r$, all other coordinates and time being held fixed. In general, Q_r is made up of a combination of the individual forces and torques acting on the system. In most actual problems, expressions for generalized forces are quite easy to obtain.

Consider now a system that is acted upon by "inherent" forces such as those of gravity, springs and friction, together with various controlled driving forces. For this system the generalized forces can be expressed as

$$Q_r=F_{q_r}'+F_{q_r}, \quad r=1, 2, \cdots, n \tag{4}$$

where F_{q_r}' is the part of the generalized force due to the inherent forces, and F_{q_r} that due to the driving forces. Then Eq. (2) can be written as

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_r}\right)-\frac{\partial T}{\partial q_r}-F_{q_r}'=F_{q_r}; \tag{5}$$

F_{q_r}' can easily be obtained from the relation $\delta W_{q_r}'=F_{q_r}'\delta q_r$ where $\delta W_{q_r}'$ is the work done by inherent forces when q_r alone changes to the extent of $+\delta q_r$. In applied problems, F_{q_r}' usually will be a function of coordinates, velocity and time or of some combination of these quantities.

[1] E. T. Whittaker, *Analytical dynamics* (Cambridge Univ. Press), pp. 34–40.

Equation (5) forms the basis of this paper. To understand how it is applied, let it be assumed that the motion of each coordinate has been prescribed. Mathematically this means that Eqs. (1) are stated explicitly. Hence, by differentiation, each $\dot{q}$ and d^2q/dt^2 can be written as a function of time. Therefore, by substitution, the left-hand member of each of the n equations represented by (5) can be expressed in terms of time. Now suppose that the manner in which the driving forces $f_1, f_2, \cdots, f_n$ are to be applied has been prescribed. Then explicit expressions for each of the generalized forces F_{q_r} can be obtained at once from the relation $\delta W_{q_r}=F_{q_r}\delta q_r$, where δW_{q_r} is the work done by driving forces (introduced as unknown quantities) when q_r alone is changed by the amount of $+\delta q_r$. Introduction of these expressions for F_{q_r} into Eqs. (5) gives a set of n algebraic equations involving only the driving forces and time. A solution of these simultaneous equations yields each driving force as a constant or a function of time. As will be seen later, it is frequently possible to express driving forces as functions of generalized coordinates only.

If the inherent forces are all conservative, it is usually more convenient to employ the Lagrangian equations in the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_r}\right)-\frac{\partial L}{\partial q_r}=F_{q_r}. \tag{6}$$

Here $L=T-V$, where V is the potential energy of the system expressed in terms of the generalized coordinates. The left-hand member of Eq. (6) automatically takes care of F_{q_r}' in Eqs. (5).

Illustrative Examples

Two examples illustrating the fundamental principles involved as well as details of procedure will now be given.

Example 1.—Consider again the system shown in Fig. 1. The only inherent force is that of gravity, on m. Possible ways of applying driving forces to the system are indicated by f_1, f_2, f_3 (there are other ways, of course). Since the cart can move only along the horizontal track and the mass m only along the rod, the system has but two degrees of freedom. As previously stated, any one of the sets of coordinates (x_1, s), (x_1, x), (x, s)

can be used to represent the configuration of the system, and the forces in any one of the groups (f_1, f_2), (f_1, f_3), (f_2, f_3) can be evaluated so as to produce any prescribed motions of the system.

To begin with, let us assume that f_1 and f_2 are acting and that the desired motion is given by $\dot{x}=C_1$ and $\dot{s}=C_2$. Thus the problem is to find the required values of f_1 and f_2. We shall use coordinates x and s (others could be used). From Fig. 1 it is clear that

$$T=\tfrac{1}{2}M\dot{s}^2+\tfrac{1}{2}m(\dot{x}^2+\dot{y}^2), \quad V=mgy.$$

But $y=bx_1$ and $x=x_1+s$. If all coordinates except x and s are eliminated from T and V, the Lagrangian function becomes

$$L=\tfrac{1}{2}(M+b^2m)\dot{s}^2 +\tfrac{1}{2}m(1+b^2)\dot{x}^2-mb^2\dot{x}\dot{s}-mgb(x-s).$$

An application of Eq. (6) to this expression for L gives at once

$$(mb^2+M)\ddot{s}-mb^2\ddot{x}-mgb=F_s, \tag{7}$$

$$m(1+b^2)\ddot{x}-mb^2\ddot{s}+mgb=F_x. \tag{8}$$

These equations must be satisfied regardless of what the assumed motion is or how the driving forces are applied. With only f_1 and f_2 acting, it is seen that $\delta W_s=(f_1-f_2)\delta s=F_s\delta s$. Hence $F_s=f_1-f_2$, $\delta W_x=f_2\delta x=F_x\delta x$ and $F_x=f_2$. Substituting these values together with $\ddot{x}=0$ and $\ddot{s}=0$ (obtained from the assumed motion) into Eqs. (7) and (8), we get $f_1=0$ and $f_2=mgb$, which are the required forces.

The steps just outlined were followed in preparing Table I, which shows the required values of (f_1, f_2), (f_1, f_3), (f_2, f_3) for various assumed motions. All values for this table were obtained from Eqs. (7) and (8). Notice that expressions for the generalized forces F_s and F_x in terms of individual driving forces depend only on what forces are assumed to be acting and not on the assumed motion. Note also that, for the last assumed motion in Table I, each driving force is expressed in terms of s rather than t. Of course, t can be introduced merely by substituting $s=A\sin\omega t$ in the value given.

It is important to realize that exactly the same final results given in Table I can also be obtained by using any other convenient set of coordinates. For example, when (x_1, s) is used, L becomes

$$L=\tfrac{1}{2}(m+M)\dot{s}^2+\tfrac{1}{2}m(1+b^2)\dot{x}_1^2+m\dot{x}_1\dot{s}-mgbx_1,$$

from which, by Eq. (6),

$$(m+M)\ddot{s}+m\ddot{x}_1=F_s, \tag{9}$$

$$m(1+b^2)\ddot{x}_1+m\ddot{s}+mgb=F_{x_1}. \tag{10}$$

For any assumed driving forces the generalized

TABLE I. Values of individual driving forces required to produce various assumed motions in the system of Fig. 1.

Assumed motion	Forces assumed acting	Corresponding generalized forces F_s	F_x	Necessary values of individual driving forces	
$\dot{x}=c_1$, $\dot{s}=c_2$	f_1, f_2	f_1-f_2	f_2	$f_1=0$	$f_2=mgb$
	f_1, f_3	f_1	f_3	$f_1=-mgb$	$f_3=mgb$
	f_2, f_3	$-f_2$	f_2+f_3	$f_2=mgb$	$f_3=0$
$\ddot{x}=a$, $\dot{s}=c_2$	f_1, f_2	f_1-f_2	f_2	$f_1=ma$	$f_2=ma(1+b^2)+mgb$
	f_1, f_3	f_1	f_3	$f_1=-mgb-mb^2a$	$f_3=mgb+m(1+b^2)a$
	f_2, f_3	$-f_2$	f_2+f_3	$f_2=mgb+mab^2$	$f_3=ma$
$\ddot{x}_1=a$, $s=A\sin\omega t$	f_1, f_2	f_1-f_2	f_2	$f_1=-M\omega^2s+ma$	$f_2=ma(1+b^2)+mb^2\omega^2s+mgb$
Note: since $x=x_1+s$, $\ddot{x}=a-\omega^2s$	f_1, f_3	f_1	f_3	$f_1=-M\omega^2s-mb^2a-mgb-mb^2\omega^2s$	$f_3=ma(1+b^2)+m\omega^2s+mgb$
	f_2, f_3	$-f_2$	f_2+f_3	$f_2=M\omega^2s+mb^2\omega^2s+mb^2a+mgb$	$f_3=ma-M\omega^2s$

forces F_s and F_{x1} can easily be written down. They are not the same as F_s and F_x in Table I. Then for any assumed motion, Eqs. (9) and (10) at once give the required values of the driving forces. For corresponding conditions these forces will be just those given in the table.

Example 2. A more involved problem.—The system shown in Fig. 2 is of such a nature that the use of elementary methods in finding driving forces to produce prescribed motions would be tedious and involved. The horizontal shaft *ab* supports a uniform disk D to which an unbalancing particle m is attached and passes through a smooth hole in the vertical shaft *de*. The disk can be rotated about *ab* by a torque f_θ exerted by, say, a motor, the frame of which is fastened to *de*. At the same time *de* can be rotated about a vertical axis by a torque f_β applied to the crank C. Values of f_θ and f_β required to produce certain motions will be determined.

Angles θ and β are convenient coordinates. Imagine a rectangular system of axes with the origin at the intersection of *ab* and *de*, with the z axis along *de* and the x axis along *ab*. Measure θ from the xy-plane and β from the xz-plane. By writing transformation equations relating the θ, β-coordinates of m to its rectangular coordinates and eliminating $\dot{x}$, $\dot{y}$, $\dot{z}$ from $\frac{1}{2}m(\dot{x}^2+\dot{y}^2+\dot{z}^2)$, we obtain for the kinetic energy of m,

$$\tfrac{1}{2}m(r^2\dot{\theta}^2+s^2\dot{\beta}^2+r^2\dot{\beta}^2\cos^2\theta+2sr\dot{\theta}\dot{\beta}\sin\theta).$$

Expressions for the kinetic energy of the remaining parts of the system are easily obtained, and it follows without difficulty that

$$L=\tfrac{1}{2}I_1\dot{\theta}^2+\tfrac{1}{2}I_2\dot{\beta}^2 +\tfrac{1}{2}m(r^2\dot{\beta}^2\cos^2\theta+2sr\dot{\theta}\dot{\beta}\sin\theta)-mgr\sin\theta,$$

where I_1 is the moment of inertia of D (including m) about *ab*, and I_2 is the sum of ms^2 and the moment of inertia of the entire system (not including m) about *de*.

If the driving torques are applied as stated, it is obvious that $F_\theta=f_\theta$ and $F_\beta=f_\beta$. Hence an application of Eq. (6) to L gives

$$f_\theta=I_1\ddot{\theta}+msr(\mathrm{d}^2\beta/\mathrm{d}t^2)\sin\theta +mr^2\dot{\beta}^2\cos\theta\sin\theta+mgr\cos\theta,$$

$$f_\beta=I_2(\mathrm{d}^2\beta/\mathrm{d}t^2)+mr^2(\mathrm{d}^2\beta/\mathrm{d}t^2)\cos^2\theta -2mr^2\dot{\beta}\dot{\theta}\cos\theta\sin\theta+msr\ddot{\theta}\sin\theta+msr\dot{\theta}^2\cos\theta.$$

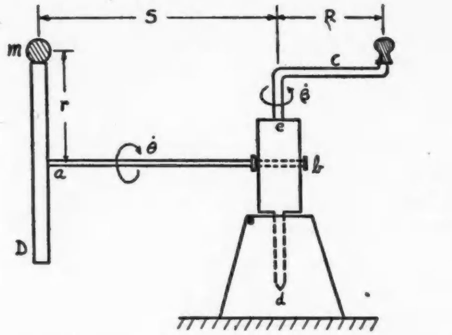


FIG. 2. Mechanism for illustrating the determinations of complex driving forces.

These are the torques required to produce any specified motions of the system whatever.

For $\theta=\theta_0=$ const. and $\beta=\beta_0=$ const., $f_\theta=mgr\cos\theta$ and $f_\beta=0$, values easily seen to be correct.

For $\theta=\omega_1t+\theta_0$ and $\beta=\beta_0$, $f_\theta=mgr\cos(\omega_1t+\theta_0)$ and $f_\beta=msr\omega_1^2\cos(\omega_1t+\theta_0)$, which can be checked by elementary considerations.

For $\theta=\omega_1t+\theta_0$ and $\beta=\omega_2t+\beta_0$,

$$f_\theta=mr^2\omega_2^2\cos\theta\sin\theta+mgr\cos\theta$$

and

$$f_\beta=-2mr^2\omega_1\omega_2\cos\theta\sin\theta+ms\omega_1^2\cos\theta.$$

These relations can, of course, be rewritten in terms of t.

For, say, $\theta=\theta_0+\omega_1t+\frac{1}{2}\alpha t^2$ and $\beta=B\sin(\omega_2t+\delta)$, values of f_θ and f_β follow at once by the same procedure, whereas to obtain them by elementary methods would be tedious and involved.

Concluding Remarks

The task of determining driving forces necessary to produce prescribed motions of a complex dynamical system may be difficult by "elementary" methods. The examples just concluded show clearly that the method here presented reduces this task to simple routine work, even for the most involved system and motions.

It is believed that this treatment will be of pedagogic interest to the instructor in Lagrangian dynamics; and, owing to its generality and ease of application, it should prove to be a valuable tool in the hands of the applied scientist.

A Journey to the Moon and Back

Henry A. Erikson
University of Minnesota, Minneapolis, Minnesota

THIS is a semitechnical examination of some of the problems which must be solved in order that a journey to the moon and back may become possible. The journey consists in leaving the earth with a velocity which will cover the distance to the moon in a reasonable period of time and arrive with a residual velocity which will permit a swing around the moon at a height of about 10 mi, then proceed back to the earth for a safe landing.

After considering various methods, including a direct take-off away from the earth when the moon and sun are in the zenith, the author finds that the plan given below is the most feasible because of its minimum power requirement, simplicity of navigation and adaptability to the technics of modern aviation.

The problems involved in the plan will be considered in the order in which they arise. First it is well to have clearly in mind the principle underlying rocket power, which is the only means available for travel in air-free space. If m is the mass of gas escaping from a rocket jet *in one second*, and v the velocity, relative to the rocket, with which the gases leave the jet, then mv is the change of momentum, in one second, of the escaping gases. But the change in momentum per second is numerically equal to the force with which the escaping gases react on the rocket. Therefore, by Newton's second law,

$$mv = (M - mt)\mathrm{d}V/\mathrm{d}t, \qquad (1)$$

where $M-mt$ is the mass of the rocket and fuel less the mass of the fuel consumed in the time interval t, and $\mathrm{d}V$ is a vanishingly small change in the velocity of the rocket relative to the earth in the correspondingly small time interval $\mathrm{d}t$. The expression $(M-mt)\mathrm{d}V/\mathrm{d}t$ is, therefore, the time-rate of change of momentum of the rocket and unused fuel at any instant. By integrating Eq. (1), assuming both m and v to be constant, we obtain

$$V = v \log_e \frac{M}{M - mT}, \qquad (2)$$

where T is the time interval, starting from rest, during which the rocket was given the velocity V relative to the earth. Equation (2) gives the characteristics of a rocket's performance. When mT becomes equal to the total initial mass of fuel carried, then $M-mT$ becomes equal to M_0, the structure-payload, and Eq. (2) becomes

$$V = v \log_e (M/M_0),$$

or

$$M/M_0 = e^{V/v},$$

where M/M_0 is the total initial mass per unit structure-payload. Since $e = 2.72$ it is apparent that, for any fixed value of the exhaust speed v, the value of M/M_0 increases very rapidly as the required speed V increases. On the other hand, for a fixed speed requirement V, the value of M/M_0 decreases rapidly as the exhaust speed v is increased.

In the plan adopted for the journey the rocket-plane, or ship, as it will be called, must have a total speed potentiality of about 4 mi/sec. If, therefore, an exhaust speed of 2 mi/sec is assumed—a speed that has not as yet been attained but is within the theoretical value of some of the fuels available—then the value of M/M_0 is 7.39; that is, for every ton of structure-payload there must be a total mass of 7.39 tons. If the structure-payload is 3 tons the total mass of the ship, including fuel, must be 22.17 tons.

If, however, an exhaust speed of 3 mi/sec were available, then M/M_0 would be 3.8; that is, a total mass of 3.8 tons would be needed for every ton of structure-payload. If the structure-payload were 2 tons, the total mass would have to be 7.6 tons.

An exhaust speed of 3 mi/sec is at present extremely improbable since it is barely within the theoretical value of the most favorable fuel now available. If, however, it becomes possible to obtain atomic energy through the use of comparatively small quantities of fissionable materials, it may be possible to vaporize materials such as water, mercury or lead and thus obtain the required speed. It must be kept in mind that the velocity of a rocket depends on the magnitude

of the escaping mass as well as on its speed of escape.

If we proceed on the basis of an exhaust speed of 3 mi/sec, then the best strategy seemingly will be to use a carrier plane from which to launch the ship at a height of 55 mi and with a horizontal speed of 4.9 mi/sec and a vertical speed of 0.82 mi/sec. The ship will then coast to a height of 100 mi where, after minor corrections in the velocity by means of the ship's own power, it will automatically move in a circular orbit about the earth at that height with a constant speed of 4.86 mi/sec. Assuming an exhaust speed of 3 mi/sec and a launching time of 200 sec, and allowing for gravity and air resistance, one finds that the carrier plane must have a total mass of about 135 tons. This mass, though large, is of the order now planned for commercial transport purposes. To meet the mass requirements of the journey on the basis of the exhaust speeds available at the present time would require a massiveness of thousands of tons.

A complete analysis may show that for the carrier plane a combination of the air-stream type of engine and the rocket type will afford the best solution and thus make the earth's atmosphere a help rather than a hindrance.

On the basis of the afore-described ship, let us now consider the details of the journey. The main power jet is in the axis of the ship at the stern. This jet is able only to increase the speed of the ship in the direction of the axis, the center of mass of the ship being also in the axis. For maneuvering purposes there are four small jets in the stern, set at right angles to the axis and to each other. By means of these four jets the axis of the ship may be given any direction, even turned so that the ship is traveling stern first. Two opposite jets of the four are also mounted so that their angles in the plane of the four jets may be changed, thus making it possible to roll the ship. To change the direction of the axis, the jet on the side of the direction desired is turned on for a brief interval, then the opposite jet is turned on for an equal interval in order to stop the rotation. The axis of the ship now has a new direction, but the ship is still moving in the original direction. By means of the power jet the ship is then given a speed increment in the new direction of the axis. The resultant of this increment and the original velocity will give the new velocity.

The magnitude of the velocity of the ship, its direction and the direction of the axis can be determined only through observation on distant objects. Within the ship there are no means by which these quantities may be determined. An increment in the speed in the direction of the axis is determined by means of an integrating accelerometer.

The ship is a monoplane of the jet-propulsion type and is designed for maneuverability and stability in air travel. The housing is of sheet aluminum alloy of sufficient strength to sustain air at atmospheric pressure within when there is a vacuum outside. The walls have horizontal compartments so that the air in the compartments exposed to the sun may be interchanged with the air within by means of a manually driven fan. There is an extra supply of compressed oxygen and also air-purifying equipment. There are aneroids, chronometers and cameras. Through plastic windows there is visibility in all directions. The cockpit is fairly spacious for a crew of two (the navigators). The 5.6 tons of fuel and exhaust mass occupy about 1800 ft.^{3} To obtain full advantage of the 0.287-mi/sec tangential speed of the surface of the earth at the equator due to the earth's rotation, it was decided to take off in the easterly direction in the plane of the equator when the moon is in the zenith and sun just east and slightly north of the zenith.

Let us consider the instant the carrier plane releases and pulls away from the ship at the 55-mi level. During the 200 sec of ascent the navigators experienced a weight of about four times their normal weight. Now, upon release from carrier, they experience no weight. The ship rises automatically until the 100-mi level is reached and, after the corrections referred to above, begins to move automatically in a circular orbit about the earth. The navigators are now living under the neutralized gravity condition, one of the unknown factors of the journey.

One of the men pushed himself upward and found that he proceeded to the ceiling. He pushed against the ceiling and returned to the seat. He raised himself above the seat and found he remained suspended in air. No moving of arms or legs would help. A swimming action gave a small result. He threw his shoe upwards; the shoe went

to the ceiling, and he returned to the seat. He was unable to walk. When, however, he placed a small magnet in each shoe he found that he could walk on the sheet-iron floor. But walking was no effort. All liquids had been placed in compressible tubes. When the water container was compressed the water collected into a transparent sphere at the tube. When the tube was jerked away the water sphere floated in air. By means of a straw it could be sucked into the mouth. If one of these spheres entered a glass, then upon contact the water collected in the bottom of the glass but the surface became wet inside and out. Turning the glass bottom side up made no difference. The situation had been anticipated by attaching magnets to all movable objects and placing them on iron shelves. When the navigators turned on the power jet the sensation of weight returned. In order to minimize mass, no cots had been provided. Now they did not need any. It was fully as comfortable to sleep standing up. The navigators seemed to become accustomed to their strange life. Seeing, hearing, feeling, sleeping, thinking, digesting—all seemed quite normal.

It is of interest in this connection to note that there seem to be no functions in the human body, such as assimilation, heart action and sense perception, that depend on gravity. The basic processes are intermolecular and involve surface tension, osmosis, chemical action and electric effects, all of which are independent of gravity. It should be noted that within the ship the gravitational forces between bodies and particles are normal. Air pressure is, of course, vital, but it can be produced by means other than gravity. The human body seemingly has adapted itself to the earth's force of gravity, but this does not necessarily preclude its ability to function independently of gravity. The outstanding result of gravity is preventing matter from disseminating throughout space.

Not having had the benefit of previous tests, the navigators observe closely the effect of this new condition within the ship. They are now in a moving laboratory quite ideal for the study of various effects, including cosmic phenomena. Should any serious effect develop they could return to earth; the ship's power and design would enable them to do so. The experience gained would have justified the effort.

The ship is now exposed to high energy cosmic-ray particles and short-wave photons. What insidious effect may arise from these is unknown. The instruments for the detection of these rays are observed closely. The ship is also exposed to intense solar radiation, and direct exposure to it would be serious. Fortunately, the effect of this radiation is minimized by the absorbing material of the enclosure. This laboratory would be quite ideal for the study of variation in the solar energy.

The ship, as stated, is now in its circular orbit about the earth at 100-mi level. Below through the telescope the navigators get glimpses of the earth's surface between hazy areas. After 20 min they approach the edge of the earth's night. Now the night's edge is passing beneath them. The sun is approaching the horizon. Now they pass into the earth's shadow. By means of radar they check the constancy of their altitude. Any change must be corrected by means of the ship's own power. The thermometers indicate a falling temperature. They are observed carefully; an excessive drop would mean having to seek a lower level. In about 30 min indications of a dawn begin to appear ahead. The dark night prevails below and behind, moderated by a faint haze. Above, the stars stand out as brilliant gems against a black background. Now the dawn is about them, and on the horizon ahead they get a glimpse of the sun's edge. It is sunrise. Now the sun is above the horizon. They are again in the light of day. The night's edge is passing below. They see the moon's cresent adjacent to the sun—the moon that is their objective. The temperature is rising slowly. Now after about 1.5 hr they are above the starting point. By radio they report to their base and relate their experience and progress. They have completed their first circle about the earth. The plan is to circle the earth 2.5 times before taking the next step.

The problem now is to determine their speed accurately. This they do by recording the time interval between two successive vanishings of the sun's edge at sunset, also between two appearances of the sun's edge at sunrise, an interval of about 1.5 hr. The radar gives the altitude. The time and distance of a circuit being known, they compute their speed.

At their 100-mi altitude what is left of the atmosphere must be primarily hydrogen. Its

density is of the order 10^{-8} lb/ft^3. This means that in 1 hr the ship will encounter about 30 lb of gas per square yard cross section. This will retard or heat the ship only by a neglible amount. An automobile traveling at the rate of 60 mi/hr encounters about 115 tons of air per square yard cross section in 1 hr. The navigators have now completed the second circuit around the earth. As they have noticed no ill effects and as the preparations seem adequate they decide on a take-off at midnight. They check the constancy of their altitude with care; also their speed tangent to the earth's surface as well as the tangency of the ship's axis. The time length of the earth's night has been determined so that the time of midnight, when the ship will be on the earth-moon line, is known.

Now the instant of midnight is at hand. They turn on the power jet. Immediately they experience a strong sensation of weight in the reverse direction of the ship's axis. They observe the integrating accelerometer with care, for it gives the total increase in the speed. The meter now registers an increase of 1.95 mi/sec, and the power is shut off. The sensation of weight ceases and the ship is now traveling with a speed of (4.86+1.95) or 6.81 mi/sec. The ship is now on its way.

The radar records an increasing altitude. A dawn is ahead and now the rising sun is on the horizon. They enter the light of day, the perpetual day of space. Adjacent to the sun they see faintly the moon's crescent, their destination. The earth is receding rapidly. The ship is in space controlled by the earth and the moon. It responds automatically to their combined force and moves in an orbit the plane of which contains the earth and the moon.

Since the ship was launched in the plane of the earth's equator, the plane of the ship's orbit does not coincide with the plane of the moon's orbit about the earth. But this does not introduce any difficulty or defeat any purpose. The earth and the moon also move in orbits about their common center of gravity. Since the ship is part of the earth-moon system it adjusts itself automatically to these motions. For these reasons the navigators are now free of any navigation problems, except the correction of any discrepancy that may develop through errors in launching.

A master chart has been prepared that shows the speeds and their directions at all points in the ship's orbit. The navigators' task is now to check the ships performance with the master chart. The radar gives the distance to the moon at any instant. Two distances a known time interval apart give the ship's speed relative to the moon. Since the radar can also give the distance to the earth, so that the speed of the ship relative to the earth is also known, the speed of the ship and its direction can be computed and compared with the master chart. The angle between the earth and the moon at the ship is determined by means of a reflectoscope.

The life of the navigators is now a leisurely routine. They eat, sleep, and bide their time. For them day and night are no more as they are no longer a part of the earth. In all directions they see the starry heavens, a new human experience. The earth is now within the angle of vision of the eye, and is brighter. The thought of collision with meteorites comes to their minds. A computation by Kaplan indicates that in the space about the earth there is about 1 gm of meteoric material per 20 mi^3. All of this in one piece would pierce the ship. The navigators consider this a traveler's chance and dismiss the subject.

The chronometers have recorded the passing hours, and the ship is now in the midway plane, as the radar indicates equal distances to the earth and moon. The earth now appears only about eight times as large in diameter as the moon. They are each about 120,000 mi away. The ship's position checks with the chart. The instruments are recording the intensities of the cosmic radiations. Outside, the aneroid records no pressure; inside, the pressure is normal. The condition of the air is checked, the oxygen content noted and replenished and the air purified from time to time. By means of the manually operated fan air circulation is obtained, and also a fairly uniform temperature.

At about 27,000 mi from the moon, the earth and moon appear of equal size. Now the ship is in the plane at right angles to the earth-moon line, which contains the gravitational neutral point of the earth and moon. The gravitational force of the moon is equal to that of the earth. The neutral point is 23,800 mi from the center of the moon. The speed of the ship is a minimum and

now begins to increase. The gravitational force of the moon on the ship is becoming larger than that of the earth. The ship is approaching the earth-moon line. The angle between the lines from the ship to the earth and moon is approaching 180°. Now they are on the earth-moon line and crossing it with a velocity in the general direction of the limb of the moon on the crescent side. The navigators check their speed relative to the moon and compute their speed in the direction of the moon's limb. The ship must arrive above the moon with a speed such that it will cross the earth-moon line behind the moon with a speed of 1.26 mi/sec at a height of 10 mi. A careful check with the master chart is made. Any discrepancy will have to be corrected by means of the ship's own power. Apparently there is to be no need for this. This is a severe test of the precision at take-off.

The ship has now crossed the earth-moon line and is on the second loop of its figure-of-eight orbit. The moon half fills the visual angle. The off-crescent part of the moon is faintly visible owing to the sun's light reflected from the earth. The navigators now roll the ship 180° so that the moon's surface will appear below. They are approaching the eastern side of the moon's crescent as seen from the earth. The moon's surface is greater than the visual angle and the navigators are beginning to feel a part of it. Now the moon's curved edge extends across the visual area.

The edge of the moon's night is passing below them; they are entering the moon's day. Ahead is the illuminated horizon of the moon. The moon's gray cratered surface lies before them—the back of the moon, never before seen by man. The cameras are recording the passing details on the moon's surface. There are no clouds to mar the view. The aneroid shows no pressure outside. There is no evidence of an atmosphere. The temperature is rising because of the sun's radiation reflected from the moon's surface. The fan is not able to check the rise. Should the heat become excessive they would turn the axis of the ship into the vertical. Fortunately the time of transit is short. They are now again on the earth-moon line; the half circuit is completed. Their speed is 1.26 mi/sec, and the radar indicates a height of 10 mi above the moon's surface.

An interesting possibility now confronts them. By turning the ship so that it is traveling stern first they could reduce the speed, by means of the ship's power, and settle to the moon's surface. If the ship were on wheels they could, by means of its power, maneuver about and view the moon's surface through the lookouts and take off again, as their weight on the surface is only 0.17 normal. This, however, would deplete the fuel required for the homeward journey. So, instead, they release an engraved bronze cube at a high reverse speed. It settles to the moon's surface, there to remain through the ages, a silent evidence of man's visitation.

They sweep on. In all directions they have the moon's horizon. Beneath them is an unspeakably silent evidence of a cauldron of the past. Now at the horizon ahead a bright edge appears. It rises and soon the full round face of the earth is in view. It appears about eight times as large in diameter as the full moon viewed from the earth.

The radar shows that their height above the moon's surface is increasing. Now the edge of the moon's night is passing below. They are leaving the moon. Their homeward journey has begun. The ship, owing to their precision take-off, has followed the chart. The navigators are at ease as to the ship's performance on the homeward journey. They can relax and bide their time.

Their speed is now decreasing; they are moving toward the point of recrossing the earth-moon line. When they cross the line, they roll the ship 180° about its axis. They are now in the neutral plane and have minimum speed. Now they have passed it and their speed is increasing. The earth is in control.

Having ample leisure the navigators view the starry sphere about them. Off to one side is Polaris; there is the Dipper pointing the way. In the opposite direction is the Southern Cross. Over here is the star Sirius and the constellation Orion. Yonder is Venus. All seem like distant friends. Ahead is the earth pulling them homeward. The chronometers have recorded the passing hours of the homeward journey. The travelers have again experienced the perpetual day of space.

The earth now half fills the visual angle. They are moving toward the evening side, in the orbit which at midnight would bring the ship to the take-off point on the 100-mi level on the earth-moon line.

They turn the ship so that it is traveling stern first and turn on the power jet until the meter shows a reduction in speed of 1 mi/sec. They then turn the ship so that it is headed in the direction of their velocity. In this they are guided by the master chart. The mass of the ship has now been reduced over one half since take-off, owing to the discharge of mass through the jets. The ship will now enter the earth's atmosphere at a lower level.

The earth is a vast mottled gray area. Before them is the curved limb of the earth on the evening side. Through the telescope they begin to observe details. The accelerometer begins to give indications of retardation in speed. They begin to notice a sensation of weight in the direction of their velocity. They are entering the earth's atmosphere. They observe closely the thermometers which give the temperature of the forward housing.

The edge of the earth's night is passing below them. The sensation of weight is increasing. The radar shows a diminishing height above the earth's surface. They are approaching the earth's night. Now they enter the earth's shadow. The sun has set. The thermometers are indicating a rise in temperature, are now at 80°F. They gradually move the ship's controls into the upward position until the radar shows a constant altitude. The temperature becomes constant. The ship is under control. From now on it is a matter of glider technic. Now they are at midnight and below their point of take-off. Had they not reduced their speed they would here begin a second journey. The ship has speeded on in its lower course. It is now no longer in the control of interplanetary forces. It enters the dawn and then into the light of day. It is sunrise.

From the total time since take-off they determine the time of day at their base and find it to be 9 A.M. As the minutes pass their altitude is gradually diminishing with the speed. Now it is 9 A.M. They are above their base. They converse with the base and report their progress. They again enter the earth's night and later the light of day at sunrise. They again pass over their base. At midafternoon they decide to circle over the earth's daylight area. They bank their plane (the ship). On they sweep in a constantly diminishing circle about their base and also with diminishing speed and altitude. Their path is a conical spiral. Now the circle is but a few miles across. They report their intention to land. After a few more circles they swing into line with the landing lane. Now the ship touches the earth, it moves gently and comes to rest.

A 500,000-mi journey is at an end.

Microwave Optics

C. L. ANDREWS
General Electric Research Laboratory, Schenectady 5, New York
and
New York State College for Teachers, Albany, New York

MICROWAVES have for man a unique place in the electromagnetic spectrum. They are the length of a man's hand and can be measured with an ordinary meter stick. Heretofore, the study of the nature of electromagnetic waves in the elementary laboratory has been with light or with 3-m radio waves. It was as though in mechanics we could work with objects less than a hundredth of a hair's breadth in diameter or twice a man's height in diameter and no sizes between.

H. K. Schilling[1] has developed widely used sound equipment with which certain classical wave experiments, common to both sound and electromagnetic waves, can be performed with sound waves of convenient length. G. F. Hull, Jr.[2] has described equipment for the production of 20-cm electromagnetic waves and for the study of microwaves in wave guides. He also described the conversion of the 20-cm oscillator into a second-harmonic 10-cm oscillator with a 3-mw output. Those who have agonized over a coherer to detect and demonstrate the nature of Hertzian

[1] H. K. Schilling, *Am. J. Physics* (*Am. Phys. T.*) **4**, 206 (1936); **5**, 280 (1937); **6**, 156, 265 (1938).

[2] G. F. Hull, Jr., *Am. J. Physics* **13**, 384 (1945).

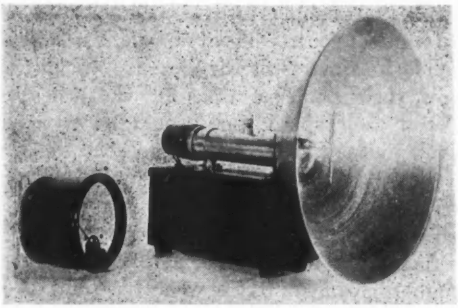


FIG. 1. Transmitter and intensity meter.

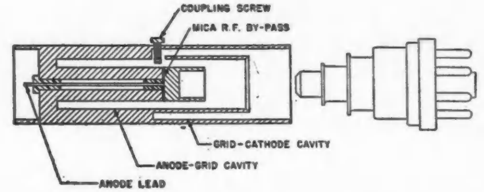


FIG. 2. Microwave oscillator employing disk-seal triode. The oscillator is 4 in. long and drawn to scale.

waves will appreciate how much simpler is the task of G. F. Hull, Jr., than was that of G. F. Hull, Sr.,[3] 50 years earlier.

The present paper is concerned with the study of microwaves in free space and the performance with microwaves of the experiments commonly discussed in textbooks and laboratory manuals in physical optics. The essential units designed for the study of microwave optics are a transmitter for production of the radiation and an intensity meter with which to explore the field of radiation. These hand-sized units are shown in Fig. 1.

The transmitter is a single unit comprising power supply, oscillator and antenna. The power supply consists of a transformer with 6.3-v filament leads, 200-v power leads and a 10,000-ohm variable resistor in the plate line with which to adjust the output power. When the oscillator is assembled, the high voltage circuit is completely shielded.

The oscillator (Fig. 2) consists of two coaxial resonant cavities with a General Electric 2C43 disk-seal triode as an integral part of the cavities. The disk-seal provides the simplest means of producing continuous microwaves.[4] The length of each of the cavities, including the portion in the tube, is three-quarters of a wavelength. The screw in the top of the oscillator is inductively coupled to the shorted end of the grid-cathode cavity and capacitively coupled to the grid-plate cavity. This general type of oscillator with two resonant cavities is described by A. M. Gurewitsch.[5] It is not a reentrant oscillator.

The microwave oscillator looks more like a whistle than a conventional radio oscillator. Indeed, for the elementary student, the graphical description of standing longitudinal waves in an organ pipe makes a good analogy for the graphical description of standing transverse electromagnetic waves in a coaxial resonant cavity. Figure 3 (a) is a diagram of a quarter-wave sound resonator, closed at one end, at the moment when the displacement to the left is maximum, and (b) shows this same tube a half-

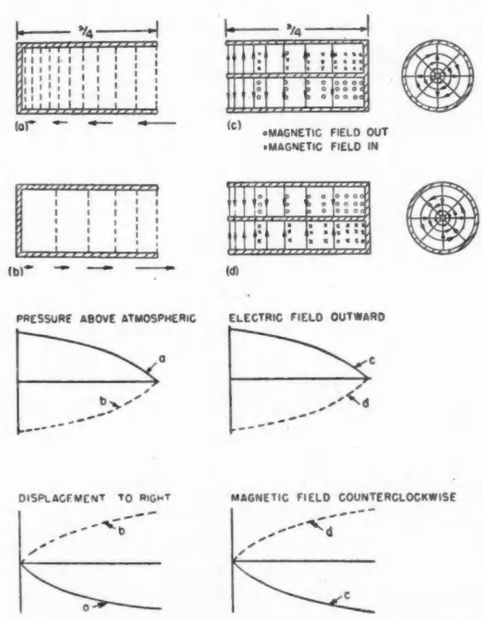


FIG. 3. Standing-wave analogy: *left*, organ pipe; *right*, microwave oscillator.

[3] G. F. Hull, Sr., *Physical Rev.* **5**, 231 (1897).

[4] The peculiar advantages of the disk-seal triode are discussed by its designer, E. D. McArthur, *Electronics* **18**, 98 (1945).

[5] A. M. Gurewitsch, *Electronics* **19**, 135 (1946), esp. Fig. 2.

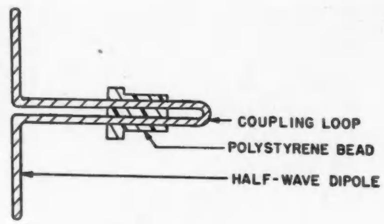


FIG. 4. Antenna.

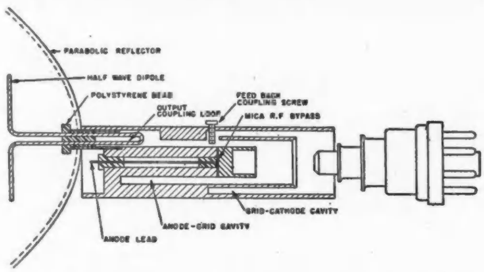


FIG. 5. Oscillator with the antenna coupled to the plate-grid cavity.

period later when displacement to the right is maximum; (*c*) is a coaxial resonator, closed at one end, when the radially outward electric field is maximum, and (*d*) is the same tube a half period later when the radially inward field is maximum. The shorted end of the coaxial electric resonator corresponds to the open end of the sound resonator.

Figure 4 is a sketch of the half-wave dipole antenna and output coupling loop. As shown in Fig. 5, this loop is coupled into the shorted end of the grid-plate cavity, where the oscillating magnetic field is the strongest.

The wavelength in air is about 12.5 cm, varying by 10 percent with change of tube. The oscillator has been made with fixed frequency both for simplicity in use and to make possible inductive output coupling, which is less critical of adjustment than capacitive coupling through the side of the oscillator. The power output is a few hundred milliwatts.

The intensity meter, Fig. 6, is a crystal detector with a half-wave dipole antenna, a short coaxial line, a sealed-in silicon crystal and a layer of Scotch cellulose tape wrapped around the crystal holder as a by-pass for the radiofrequency. The quarter-wave stub acts as a by-pass for the rectified current without reflecting the radiofrequency wave at that point. The load is a 50-μamp General Electric model DO-40 microammeter. Since the rectified current through the crystal is very nearly proportional to the square of the potential across it, the meter reading is proportional to the intensity of radiation. The metal disk a quarter wavelength behind the antenna acts as a reflector. Since the electric field strength undergoes a change in phase of 180° upon reflection, the reflected wave arrives at the antenna in phase with the incident wave.

If a 50-percent reduction in sensitivity can be tolerated, the quarter-wave stub may be omitted and a screw used as d.c. by-pass between the cylinder and center line. The best position for this short is found to be at the antenna. Thus the dipole may be one continuous rod, as shown in Fig. 7. With the quarter-wave stub omitted, the intensity meter is sensitive to a broader band of frequencies.

Either type of intensity meter is of ample sensitivity for use with the transmitter for studies of microwave optics in the laboratory or lecture room.

Geometrical optics of microwaves cannot be studied apart from physical optics. A mirror shows a diffraction pattern. A wedge of dielectric used to study refraction has interference bands across its face. If microwaves are used for hunting beams in a wall, the observer should be acquainted with the interference pattern of Lloyd's mirror in order to find the flat sides of the beam. Attention must be given to polarization. In other words, the teacher of physical optics may capitalize on all the annoyances of microwave radar.

With the transmitter and intensity meter

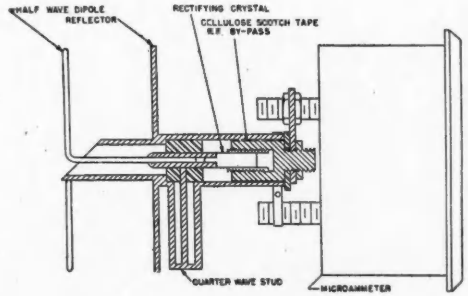


FIG. 6. Intensity meter.

described here, laboratory experiments and lecture demonstrations have been made of Young's interference from secondary sources, Lloyd's mirror, standing waves, interference in thin films, the Michelson interferometer, Fresnel diffraction, diffraction pattern of a parabolic reflector, polarization, reflection and transmission of components by parallel wire screens and elliptic polarization.

The individual experiments will be described subsequently. It is sufficient to note here that Young's interference pattern is spread across the laboratory table and that the wavelength can be determined directly from the difference in distances from an interference minimum to the two secondary sources. "Thin films" of glass or lumber for studies of interference are of thickness that can be measured on an ordinary centimeter scale. Circular apertures and rings for demonstrations of Frèsnel zone theory are the same size as the circles one draws on the blackboard. Planes of polarization from dipole antennas are readily recognized, and the parallel wires of a polarizing screen are more easily visualized in a demonstration lecture than are the principal planes of a crystal.

As an example of how simple microwave demonstrations may be, Fig. 8 is a picture of the Michelson interferometer for microwaves set up and tried out in a half hour before class. Figure 9 is a diagram of the interferometer. The movable mirror B is a flat-sided wastebasket (shown at the extreme left in Fig. 8) which could be slid between two meter sticks as guides; the fixed mirror A is a window screen; and the half reflecting mirror is a piece of chicken wire. For every half wavelength that the basket was

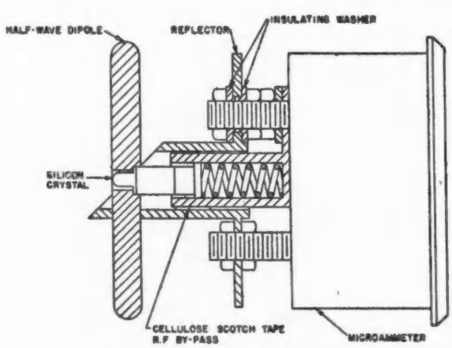


FIG. 7. Simplified intensity meter.

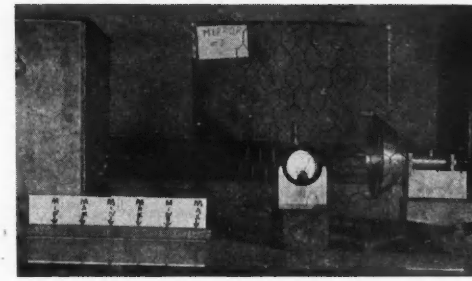


FIG. 8. Michelson interferometer for microwaves.

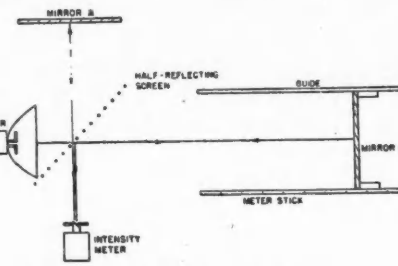


FIG. 9. Diagram of Michelson interferometer for microwaves.

moved, the meter reading dropped sharply to zero; for positions between, it rose nearly to full scale. This demonstration is not described as a brief for makeshift lecture demonstrations but as encouragement to those who would build their own demonstrations of microwave optics.

For large lecture demonstrations, audio means of detecting intensity maximums and minimums are often preferable. Since the waves are modulated at 60 c/sec, the output of the crystal detector may be fed to an audio-amplifier and speaker instead of the meter.

Although the greatest value of microwaves to the teacher of physics is in explaining the nature of electromagnetic waves in general, microwaves also deserve attention in their own right. In those colleges where the department of physics assists in the basic training of electrical engineers, if the department fails to offer a course in microwave optics for the engineers, it is likely to be taught in the College of Engineering as a branch of engineering.

It is a pleasure to acknowledge the support and advice of W. C. White and E. D. McArthur who made it possible for me to develop this microwave educational equipment.

A Simple Apparatus for Measurement of the Index of Refraction of Air

Paul S. Delaup
Southwestern Louisiana Institute, Lafayette, Louisiana

THE Raleigh refractometer[1] provides an accurate method of measuring the index of refraction of air. However, there are several practical difficulties in obtaining results with the usual form of the apparatus. Because of the large separation of the slits, the interference fringes are close together and are difficult to see and count. Also, when long tubes are used and the air removed, the fringes are no longer visible, so that it is necessary to remove only a part of the air and calculate the index by proportion.

These difficulties are overcome in the apparatus shown in Fig. 1. With the construction shown, the two slits are very close together and large, clear fringes are easily obtained. Also by using a short tube (2 cm), all the air may be removed and the fringes may still be seen.

The apparatus consists of a brass or copper tube 2 cm long, cut in half and soldered to a flat piece of brass about 0.25 mm thick. A side tube for connection to a vacuum pump is soldered to the tube. A thin piece of glass is cemented to each end with Cenco-Sealstix cement. On the outside of one of the pieces of glass is glued a piece of black paper with a slit about 1 mm wide cut in it, in such a position that the end of the brass plate bisects the slit in the paper, thereby forming two slits.

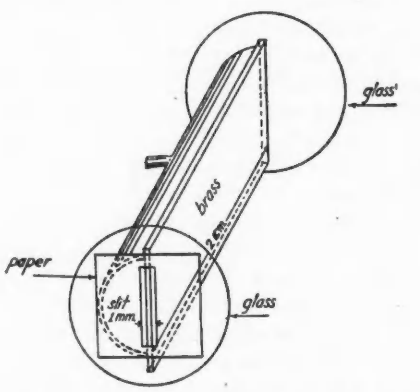


Fig. 1. Diagram of apparatus for measuring index of refraction of air.

The apparatus is set up as shown in Fig. 2. The slit end of the special apparatus E is toward the lens. A straight-filament lamp is placed about 6 ft away and a red glass filter mounted in front of it. The system can be lined up by removing the filter and forming the image of the two slits on a piece of paper. It is important that the slit be parallel to the filament if the clearest fringes are to be obtained. The focal length of the lens is about 50 cm.

The apparatus E is connected to a vacuum pump by a rubber tube. As the air is removed the fringes will shift. After as good a vacuum as possible is obtained, the pump is disconnected, air is slowly admitted by adjusting a pinch clamp on the rubber tube, and the number of fringes passing the cross hair in the eyepiece is counted. Fractions of a fringe may be estimated. If difficulty is found in admitting the air slowly enough a large bottle may be placed in the vacuum system.

The index of refraction μ can be calculated by means of the equation

$$\mu = 1 + (\lambda N/l), \tag{1}$$

where N is the number of fringes shifted, l is the length of path, λ' is the wavelength in air, and λ is the wavelength in vacuum. To obtain Eq. (1),

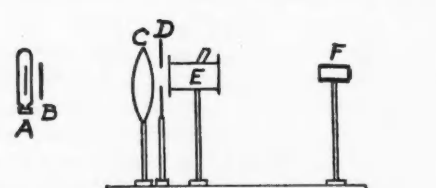


Fig. 2. Arrangement of apparatus: A, straight-filament lamp; B, red glass filter; C, lens, about 50-cm focal length; D, screen; E, apparatus of Fig. 1; F, eyepiece with cross hairs.

[1] J. K. Robertson, *Introduction to physical optics* (Van Nostrand, 1935), p. 199; Wood, *Physical optics* (Macmillan, 1934), p. 173.

note that N is the difference between the number of waves in the air and vacuum paths; hence

$$N=\frac{l}{\lambda'}-\frac{l}{\lambda}=\frac{l}{\lambda}\left(\frac{\lambda}{\lambda'}-1\right)=\frac{l}{\lambda}(\mu-1).$$

The wavelength can be taken as the average of the wavelength range transmitted by the filter. More accurate results may be obtained by replacing the lamp and filter by a strong source of monochromatic light.

Samples *Versus* Survey in Physics Courses for Liberal Arts Students

Eric M. Rogers
Princeton University, Princeton, New Jersey

THIS is a plea for the good name of physics in general education. Many changes in elementary courses are being discussed. There is talk of a thorough two-year course for "specialists"—future physicists, engineers, chemists, perhaps mathematicians; and I expect such courses will come, though probably disguised as a group of courses. But I want to raise here some doubts and questions about physics courses for liberal arts students—economists, lawyers, poets—who will take only one course of physics, perhaps as their only science course in college. Such students need a self-contained course that is, in a sense, an end in itself, a course that will give them an understanding of the nature of science, with values which may spread soon to other fields of study and last long after college to help them to live in the civilized world of today.

Yet at present these students have to take either a technical course, full of groundwork for later courses, or a survey course which surveys too much too fast. Neither course gives the understanding and inspiration we hope for. A formal technical course is apt to give such students a frustrated feeling of headache rather than inspiration. They find it not only hard but pointless, since quantities, concepts and relationships are introduced and then barely used. We meet the results of this among nonscientific friends of our own generation. If we ask lawyers, writers or business men about their physics, how often do we get replies like these: "A fine course that gave real insight," or "Hard; but it made sense and I enjoyed it"? We usually hear: "I did not like it," or "I got the problems right but I never really understood it." Such replies make me feel that poor justice has been done, in teaching, to the good name of physics. No bishop would advocate for a general course in Bible-study a detailed course of textual criticism. No medical man devising a general hygiene course would embark on a detailed naming of all the bones of the body. Yet in physics we still try to do something of the kind. As a remedy, some colleges give a "survey course" which covers a huge range of topics with a wealth of fact but little rigor. Such crowded courses produce another feeling of headache by presenting too many details, with too little explanation and reasoning, or worse still, they encourage the belief that physics is a body of occult knowledge with mysteries which the omniscient teacher unfolds only partially. These pictures, of a science as a pile of unrelated facts and of the scientist as "the man who knows," give students a false idea of science. I wonder whether a survey that "covers every important topic" does contribute much to the real education of the liberal arts student.

What is needed is a course in which there is time for careful study (in classroom and laboratory) to show what physics is driving at, how it proceeds and what kind of structure of knowledge it builds. A course of "sample" topics can show how scientists tackle problems; how experiments are devised, carried out and analyzed; how theories (or hypotheses) are suggested, tested and used; and how both practical applications and further understanding grow from all this. From such a course can come, above all in the laboratory, a knowledge of the way in which we build on a firm foundation of experiment, and a recognition of the rigid honesty with which scientists abide by the results of experiment rather than by what they expect. With such aims we need to prune our syllabus, but with such aims we shall not find it hard to omit topics that are

mainly used to prepare for later courses. We can then choose topics, and means of presenting them, that will further these aims. We shall have time to expand these topics—to do thoroughly what we *do* do—and we shall use them both as examples of approach and method and as parts of a general framework. Though we cannot avoid some patchiness if we reduce the content of our course, any wise attempt to show what physics is like must, for that very purpose, retain some continuity of structure. In fact, if careful discussion of a few topics gives him a sense of clear understanding, the student will have a better chance of seeing that there is continuity of structure.

So I suggest a great reduction of content, say to half the present number of topics, not just a paring down by 10 percent. However many important topics we omit, the ones we keep will gain enough to justify the omissions. Physics is not easy, and most students need far more time and discussion than we now give if they are to grasp the meaning of a piece of physics. They may repeat formal statements and solve numerical problems correctly and yet have little understanding of the underlying principles and ideas and methods. A ruthless cutting down of syllabus seems essential for giving more valuable and lasting understanding.

Though I want to insist on limiting the course to fewer topics, I do not want to list certain topics as the "right" ones to be taught. I am sure the teacher should make his own choice, which should be influenced by the interests and abilities of both teacher and students. Here it is clearly not so much what is taught but how it is taught, and why, that will matter. So I offer only general comments on selection, with a few examples. We can gain some guidance if we think of our students ten years after college and ask what values we hope they will retain from their physics courses. With this question in mind, I think we shall want topics where the experimental approach is shown clearly, some also showing the use of hypothesis and the place of law and theory, some with useful applications to ordinary life, a few for their essential information value, all as parts of the structure of a science. Most of our choices will be parts of the present physics courses, treated with a different aim. But we may omit discussions of quantities or concepts that have no essential application in the course. For example, we may omit specific heat and its measurement. Without specific heat we can still study latent heat or transfer of heat or even thermodynamics—all of them topics of greater importance with chances to inspire. In mechanics, we may omit much of statics, which at present, because it comes at the beginning of the book, gets too big a share of time even in a technical course!

In our choice and treatment of topics, we may try to correct some prejudices that students often have, such as a faith in labels. If I ask, "Why does this stone fall?" many students reply "gravity," and regard this, or the longer word "gravitation," as an explanation in itself. We may argue them out of this faith, but they are likely to take refuge behind names again in other topics. For example, some students explain the action of a thermos flask by saying that it lessens the loss of heat by convection, conduction and radiation. Evaporation, which is quite as important, is seldom mentioned. If they learn to link this behavior with other observations, they are learning wisely, but if they put their trust in the actual names, they are indulging in elementary botany. Despite the claims of many a general-science teacher, these names, even backed by some knowledge, do not make the student a better citizen by showing him practical applications of physics. The two practical facts worth knowing about a thermos flask are (i) that the inside part is fragile and (ii) that when it is broken a new inside part can be bought cheaply in drug stores—but these facts have no proper place in a physics course. A discussion of labels and "explanations" should crop up again and again in our teaching, to teach something of the nature of science.

As examples of treatment, I want to discuss two useful topics, both of which I like to use, though neither is an essential topic that I would urge others to use. The first is *surface tension.* If we propose to spend a short time on this, say one lecture, it is just one topic among many, probably better omitted. But if we spend several hours on it, with some laboratory work, we can exhibit a microscosm of scientific work, with a marvelous variety of ramifications. Out of a study of shapes of drops, classified by angle of contact, come applications of waterproofing and wetting agents

(with useful warnings about raincoats) and hints about molecular forces which the student may use, if he is so minded, to "explain" what is happening. Experiments with thin oil films yield estimates of molecule size that can lead off on a trail of organic chemistry. Simple discussion of surface forces gives hints of uses in biology. This may lead to a discussion of competition between surface forces and volume forces, with examples ranging from the curvature of comets' tails to the strength of elephants' legs. Surface tension involves no complicated theory (in such a course), but the student meets a region of knowledge with practical and theoretical aspects that are likely to catch his fancy somewhere and give him an idea of the building up of knowledge. If he does some simple laboratory work on surface tension, he will come to think of the experimental physicist not as a wizard in his tower but as a careful, cunning worker—a Langmuir in his shirtsleeves, cleaning apparatus scrupulously, measuring carefully.

Another example, to show how we use theories: Young students tend to think of a theory as coming from nowhere, as part of the scientist's mystic knowledge. Even a sensible undergraduate retains a strong element of this view in his prejudices. To show how theory or hypothesis is really built up, I like to use the "molecular" theory of magnetism. A few simple experiments suggest the idea of molecular magnets—or rather domains. (Many a student would be content to stop short at this stage with the theory merely explaining the facts that suggested it!) Emphasizing the scientist's use of theory to suggest further advances, we make use of our theory to suggest answers to a number of questions, such as the following: (a) is there a limit to the magnetization of a steel bar?; (b) if bars of steel and of soft iron are magnetized forward and backward in an alternating field, which will grow hot faster?

The answers to both questions are of great interest to the electrical engineer. To the general student their main interest lies in the way theory suggests answers. We insist that once the theory has suggested answers, its job in the matter is done, and the next move is to test the suggestions experimentally. Successful tests encourage us to keep the theory for further use. However, we have here a rare case where theory goes further and enables us to understand something that would otherwise be meaningless. We say, "Suppose a man thinks he has magnetized a steel ring, but can find no poles. Is it possible that, in any reasonable sense of the word, the ring *is* magnetized?" Without the theory, the sensible answer is, "No; there are no poles and no magnetic moment, therefore there is no magnetization." With the theory, students suggest with delight that the domains may be arranged with circular magnetization, head to tail round the ring. They even suggest a test of this by cutting the ring—and in this view they have found something of great practical importance in modern transformers.

Here we have theory rising to a new height of usefulness, a new tool giving greater understanding. And we can say to the student, as to the Elephant's Child, "You couldn't do that with a mere-smear nose." In this I rank the molecular theory of magnetism with the use of structural formulas in organic chemistry. Without these latter, who could say how many dichlorbenzenes to expect? With a sketch of the benzene ring, we predict at once three distinct compounds.

The student's own laboratory work can and should play an essential part. For some, it is the only way in which a real understanding of the experimental nature of science is gained. So we should remove the dull cookbook treatment of experiments. Experiments should be real enquiries. The standard example is a full investigation of pendulum motion. This can be a thrilling piece of work, but we must be frank and tell the student that he is repeating an old investigation, partly for its results, partly as an example of method of attack. We should omit many routine measurements. Again I suggest omitting the measurement of specific heat by mixtures. It is dull, encourages the student to seek "the answer the book gives," and involves methods that are out of date. Almost all serious calorimetry in this century has been done by electrical methods.

Returning to my general plea, I plead for a course of samples, treated carefully to show the nature of physics and how the physicist goes about his business. I believe this would best serve our aims in education. Just as a course in classics may try to give, through literature, a vision of a civilization, so in our physics course we should try to give, through an understanding of the nature of physics, a real share of mankind's scientific culture.

Physics in General Education: Foreword

TEACHERS of college physics undoubtedly desire to participate in the current improvement of general education. Apparently physics instruction must move in two directions simultaneously, one toward a more adequate contribution to general education and the other toward giving adequate training to those having specialized interests in science. There is no immediate single direction that can be followed. It is hoped that the following four papers and discussion[1] at the 1946 Colloquium of College Physicists, State University of Iowa, will encourage experiments in an improved adaptation of college physics to general education and that these will lead to further consideration at the 1947 session of the Colloquium.—G. W. STEWART.

The Requisites of General Education

RUSSELL M. COOPER
University of Minnesota, Minneapolis 14, Minnesota

IT is no accident that thoughtful people everywhere are discussing general education. The record of recent years affords abundant proof that the American schools and colleges have been turning out graduates with amazing technical capacities, but with sad deficiencies in the art of personal and social living. Through inability to understand and control social affairs, the nation has stumbled into two wars in a generation, suffered through a severe economic depression, and now is threatened by another boom and collapse—not to mention the possibility of a suicidal atomic war. In their personal lives, the inability of people to find a normal, wholesome adjustment is reflected in the alarming increase in divorce rates, drunkenness and general disillusionment. Clearly technical education is not enough. People must also be taught to live.

It is significant that through the years specialized education has become increasingly systematic and efficient while preparation for living has usually been haphazard and fragmentary. In order to insure some breadth of education, the student has simply been encouraged to sample courses from among the hundreds of offerings in various fields, offerings that often were completely unrelated to one another and that were primarily designed not for his general educational needs but for the purposes of prospective departmental majors. For general education to really count, the student must gain a perspective and an awareness of the interrelationship among the courses pursued instead of a meaningless smattering.

Moreover, courses must be devised to meet the explicit interests of nonmajors who are seeking aid for practical living, and the teaching procedures must be carefully devised to assure those ends. In other words, the careful planning and the thorough systematic teaching that have come to characterize specialized education at its best must now be applied to general education. Only by such intensive effort can education retrieve the losses of recent years and prepare oncoming generations of students actually to solve the complex problems of our time.

Two Approaches to General Education

In order to meet this challenge to general education, institutions have developed many kinds of programs. Indeed, in the present stage of development, there are scarcely two institutions anywhere that have exactly the same approach to the problem. Nevertheless, despite these variations, the plans can usually be grouped around one of two philosophies. These two methods of approach need to be distinguished carefully, and each faculty must decide for itself which philosophy best coincides with its own basic objectives.

Colleges following one philosophy are concerned primarily with the transmission of the cultural heritage to the oncoming generations of

[1] The first two papers—by Dean Cooper and Professor Roller—appear in this issue. The remaining papers—by Professors Lloyd W. Taylor and C. N. Wall—together with a condensed version of the discussion that followed the four papers will appear in the January-February, 1947, issue.—EDITORS.

students. It is believed that if these youths can become acquainted with the great truths embodied in the various fields of knowledge and can read the thoughts of the great minds of all times, they will themselves derive the breadth of view and the wisdom which will enable them to make intelligent decisions in their own lives. The selection of truths to be imparted may vary according to the temperament of the faculty. At St. John's College the emphasis is upon the 100 great books; at the University of Chicago there is much greater inclusion of contemporary social and scientific thought; in the famous "Harvard Plan" there is emphasis upon both the cultural heritage and the contemporary. But in every instance the central purpose is to organize the systems of knowledge into comprehensive units and to impart these truths to the students for their edification.

The second approach to general education may include some of the same materials as the former, but its method of attack is exactly the opposite. Instead of looking to the past to discover what great truths can be imparted, it looks to the future to determine what are the needs of young adults in modern society. In the light of these needs the curriculum then endeavors to afford the student practical experience in problem solving, using all of the relevant data which may be procured from any source, past or present, classical or transitory. Exponents of this school of thought insist that students will learn to make sound judgments only when confronted with practical problems requiring such judgments. Curriculum content is subordinated in importance to the practical experience gained from becoming acquainted with reliable sources of data, analyzing them and weighing one factor against another. In the course of these investigations, the student may be impelled to read a considerable amount of classical material for the insights to be found there, and he almost certainly will wish to consult the best contemporary sources of information, but the process of thinking will be essentially inductive rather than deductive and will grow directly out of the practical problems of adult living.

This second plan of general education has two important advantages over the first. It is much stronger in its motivational appeal to the student because it is more closely related to practical life. The student enters upon his studies with a spirit of adventure in order to solve problems which he recognizes as vitally important. In the former plan the student is asked to read books because they are "profound" and "classical," but their importance and relevance for his own basic concerns are frequently obscure and incidental. Another advantage of this second plan of general education lies in the practical experience which the student gains in making judgments. It is one thing to see how Plato, or Galileo, or Darwin has analyzed and solved a problem, but it is quite another matter to develop the habits of systematic thinking in oneself. There is an increasing amount of evidence to indicate that systematic thinking is developed best through frequent experiences in actual problem solving rather than by admiring the mental prowess of others.

The requisites of general education that are listed below are obviously related to the second approach just described. They suggest some of the basic needs of youth that must be satisfied if students are adequately to meet the responsibilities of tomorrow.

Twelve Requisites of General Education

There is nothing sacred about the number twelve. Other persons making a list of the general education needs of young adults would undoubtedly group then somewhat differently and perhaps more logically, but the items listed below probably cover most of the essential points. The list is based upon the assumption that general education, while it involves acquiring basic information in broad areas of human experience, also must include mastery of fundamental skills, constructive attitudes and quickened appreciation. It must prepare a man not only to understand himself and the world but also to use these insights effectively in practical situations.

Among the general education needs that are common to all people we would, therefore, emphasize the following:

1. *Effective communication.* A man should be able to organize his ideas systematically and to so express them in speech and writing so that he will achieve the desired

response in others. Similarly he must develop habits of discriminating and efficient reading and listening.

2. *Responsible citizenship.* He should be prepared to participate as an active, responsible and informed citizen in the discussion and solution of the social, economic and political problems of American and international affairs.

3. *Scientific understanding.* He should understand the basic principles of the physical and biological world, their implications for human welfare and their influence on the development of thought and institutions.

4. *Health.* Since physical well-being is so important for happiness, he should know how to improve and maintain his own health and to make intelligent decisions about community health problems.

5. *Family life.* He should acquire the knowledge and attitudes that will prepare him for successful marriage and home responsibilities.

6. *Esthetic appreciation.* He should come to understand and enjoy literature, art, music and other cultural activities as an expression of personal and social experience, frequently through active participation.

7. *Personal adjustment.* He should attain a balanced social and emotional adjustment through an understanding of human behavior, the enjoyment of social relationships and the experience of cooperating with others.

8. *Philosophy of life.* He should develop a set of principles for the direction of personal and societal behavior through the recognition and critical examination of values.

9. *Vocational choice.* He should choose a socially useful and personally satisfying vocation through an analysis of his particular interests and abilities and their relation to occupational opportunities.

10. *Critical judgment.* He should develop in all activities the habit of systematic thinking and a readiness to face facts objectively and courageously.

11. *Intellectual curiosity.* He should gain a steady quickening of interest and inquiry and an eagerness to push an idea through to its ultimate conclusion.

12. *Creative imagination.* He should develop the habit of thinking for himself, searching out new hypotheses and exercising his inventive ingenuity.

The Responsibility of Physicists

In examining the foregoing list of general education needs, it will be noted that the first nine refer to aspects of living while the remaining three are concerned primarily with the sharpening of intellectual powers. Among the first nine requisites, physicists will probably contribute most to the third—the development of scientific understanding—although some important contributions can and should be made to all the other objectives. Even in the development of scientific understanding, the kind of physics course required is very different from that usually provided for prospective majors in the field. For the nonmajors it is important to give a broad view of the physical sciences and an understanding of how the basic principles relate to the thought and institutions of our age. Technical definitions and formulas that are of value only to those students who are expecting to major in the field should be largely reserved for the major courses and should not clutter up this general education offering—especially since they will soon be forgotten through nonapplication. Many institutions have developed for the general student an integrated physical science course embracing some physics, chemistry, geology and astronomy, and this trend holds considerable promise for the future.

It is probable that the physical sciences can make an important contribution to the final three requisites mentioned: the development of critical judgment, intellectual curiosity and creative imagination. It is by no means certain, however, that these values are an inevitable product of the physical science course. Some scientists have charged that many courses, especially at the elementary level, are essentially fact-giving and that even the laboratories are often largely manipulative exercises. Such courses may require careful computation, but often there is little demand for critical judgment, intellectual curiosity or creative imagination. If these virtues are to be cultivated, problems must be defined and procedures devised for achieving these precise ends. The scientific method must be carefully analyzed and then applied over and over again until it becomes for the student a habitual method of solving problems.

Even when these intellectual values are cultivated in dealing with physical materials there is no ready assurance that the student will apply the same mental qualities in the solution of personal and social problems. Transfer of such mental abilities does not occur automatically. It comes only when the manifold possibilities of transfer are carefully discussed by the professor and student at the time of the original learning, and it requires frequent exercise under professorial encouragement and direction.

Unless there is continued emphasis upon the possibilities of using the scientific method in the other aspects of life, the intellectual values

derived from the study of physical science may be largely lost, for by definition a student who is taking such a course for general education purposes does not intend to continue in the field and will not frequently be called upon to solve physical problems. Whatever intellectual maturity he may develop will be expressed almost entirely in other fields—hence the importance of teaching with these transfer possibilities continually in mind. This transfer factor provides another argument for the broader natural science course which provides many opportunities for demonstrating the widespread application of the scientific method.

It is obvious from this discussion that general education cannot be the sole responsibility of any single department in the curriculum. Every field of learning has its contribution to make, and physical science can render some of the most important services of all. The future of general education in America, and in a very real sense the future of America itself, requires that every professor recognize clearly his responsibility and reorganize his teaching to insure its fulfilment.

Some Essential Features and Uncommon Objectives of a Physical Science Course for the General Student

Duane Roller
Wabash College, Crawfordsville, Indiana

IF physical science has failed to play an important role in general education, is it not mainly because, to use Dean Cooper's words,[1] we have failed to apply to general education the careful planning and the thorough, systematic teaching that have come to characterize specialized education at its best? Old conceptions of the process of general education not only are vague but also appear to be outmoded. Curiously, there is some advantage in coming to regard this process as a form of specialized education, but one in which the primary task is to help people acquire understanding and skill in the art of living.

Some Essential Features of the Course

When the problem is approached in this fashion, certain essential features of a general course in physical science seem to become evident.

(1) Any course intended for purposes of general education should be designed specifically to meet those purposes, and no others. In particular, no consideration whatever should be given to the special needs of students who plan to major in the field covered by the course.

(2) For the culture that we are now entering, physical science is an indispensable instrument of general education, which means that it should be taught from the kindergarten up. But because physical science has numerous highly sophisticated and difficult aspects thought to be of great value to able people, a course on the more mature level of the college also seems essential. It is not unreasonable to insist that every college student be required to take such a course, of one-year duration and involving six to eight semester hours of credit.

(3) In the present state of development of the sciences, physics should supply the core material of a physical science course that will best meet the requirements of able, mature students. This is not only because physics is fundamental, but because, of all the sciences, it has the most highly developed theoretical structure and the most explicit methodology.

(4) The physical science course should in character be *selective* rather than comprehensive; and mainly *analytic* rather than descriptive. Certainly there is no need today for a course that merely extols the wonders of science. Nor is a college course that "covers the field" in a descriptive fashion likely to be effective in imparting the attitudes, understandings, modes of thought and procedures that physical science is in an almost unique position to place at the disposal of able, mature students.

(5) Individual laboratory work, genuine in character and designed especially for mature nonscience students, should form an integral part

[1] See the preceding article, p. 387.

of the course. Because of the large expense when many students are involved, college administrators have the right to ask whether there are adequate substitutes for individual laboratory work; and, too often, the answer should be "yes," that a good deal of laboratory work, *as now given* in the general course, is not much better than demonstrations, exhibits and similar visual aids. Able students who have looked forward to laboratory work with pleasure often find that their enthusiasm wanes when they actually begin it; their initial interest in conducting experiments in what they would like to regard as the workroom of the scientist soon degenerates into mechanical performance to meet course requirements. Since no place other than the laboratory offers more opportunities for stimulation of curiosity, for self-expression and for creative activity, the conclusion seems to be that most so-called laboratories for the general student really are not laboratories.

H. I. Schlesinger[2] points out that three of the main purposes of a general education are: (i) to train students in clear thinking, (ii) to teach them how or what to see, (iii) to give training in the art of translating observation and thought into well-considered action. Concerning this last purpose, he adds that most students are preparing for careers crowded with activity, yet the guidance of this strong impulse has been left almost entirely to opportunities created by the students themselves through student activities.

Laboratory work can be made a powerful tool in achieving these aims of general education—the synthesis of observation, reflection and action into a pattern of behavior; and, as Schlesinger says, if these larger purposes were made the chief aims, it would never be supposed that lecture-demonstrations could be anything more than a poor substitute for individual laboratory work.

Specific suggestions for planning and operating a cultural laboratory will be found in various articles and digests that have appeared in this journal.[3] One tendency that obviously should be avoided is to select and design the student experiments first, and then to rationalize their purpose. Almost any subject matter will contribute something to a general education; but the question should be, how much is it likely to contribute? The time one has with these students is short.

(6) The teacher of the general course must be one who is willing to make its cultural aims the only aims; moreover, he must be a good scientist of broad education if he is to contribute most to able students, many of whom possess aptitudes for their own chosen field comparable to those of the scientist for science. In a day when society is faced with the problem of sheer survival, it should be clear that general education is a major function of undergraduate science departments, needing the earnest attention of the best minds among staff members. This should be especially clear at a time when many scientists are finding it necessary to spend much time and energy on activities of a political nature. If it is not, then it would seem that we have little faith in the efficacy of education, or else that politically our concern is with the future safety, not so much of society, as of science.

(7) Tests and examinations given to general students should be specially designed to assist in judging the efficiency of the course; that is, in finding out how well the course accomplishes its aims.[4] If the test items are those ordinarily given to physical science majors, the course will soon revert to the traditional type.

Some Uncommon Objectives

Of the 12 requisites of general education outlined by Dean Cooper, the third one—scientific understanding—naturally comes first in any science course. Doubtless agreement would also be rather general that physical science can contribute materially to the last four requirements—vocational choice (which today might well be extended to include vocational flexibility), critical judgment, intellectual curiosity and creative imagination—the last three being especially significant in physical science since there they can be cultivated in situations that are relatively simple and clear.

[2] H. I. Schlesinger, "The contribution of laboratory work to general education," *J. Chem. Ed.* **12**, 524 (1935); digest, *Am. J. Physics* **4**, 55 (1936).

[3] Among them: A. A. Bless, **1**, 88 (1933); G. B. Welch, **3**, 69 (1935); M. H. Trytten, **3**, 192 (1935); H. I. Schlesinger, **4**, 55 (1936); V. E. Eaton, **5**, 47 (1937); C. J. Overbeck, **6**, 141 (1938); D. Roller, **6**, 251 (1938); J. A. Eldridge, **7**, 69 (1939); R. E. Berger, **7**, 398 (1939); M. Kostick, **8**, 331 (1940); J. S. Rinehart, **9**, 218 (1941); L. R. Weber, **10**, 58 (1942); W. B. Thomas, **12**, 53 (1944).

[4] See, for instance, Smith, Tyler and Heil, "Evaluation of student achievement in the physical sciences," *Am. J. Physics* **5**, 102 (1937).

Each of the liberal arts fields naturally should concentrate on those requisites to which it can make the most definite and effective contributions. However, it should come to be more generally appreciated among educators that the requisites just mentioned are not the only ones to which physical science can contribute in ways that are almost unique.

Effective communication of ideas.—The field to be drawn on here is rich, for the physical scientist has made important contributions to linguistics.[5] As Leonard Bloomfield says:[6]

> A typical act of science might consist of the following steps: observation, report of observations, statement of hypotheses, calculation, prediction, testing of predictions by further observations. All but the first and last of these are acts of speech. Moreover, the accumulation of scientific results (the body of "science") consists of records of speech utterance, such as tables of observed data, a repertoire of predictions, and formulas for convenient calculation.
>
> The use of language in science is specialized and peculiar . . .: the scientist's use of language is strangely effective and powerful. Along with systematic observation, it is this peculiar use of language which distinguishes science from nonscientific behavior.

But if the general student, in his brief contact with physical science, is to improve materially in his ability to use language, he must be made consciously aware of the special and peculiar attributes of scientific language and of how they make for improvement not only in communication but in the very process of thinking. His textbook should be faultless, and he should be shown how to use it efficiently. In written work he should be expected to meet standards similar to those maintained in courses in English composition, and in the laboratory he should have practice in making real reports (not school reports of the traditional kind) and in planning and executing readily interpretable diagrams, graphs and tables. Most important, he should be constantly reminded of how these various devices can be employed to advantage in extrascientific fields whenever factual knowledge is to be communicated.

A student can spend the year learning the definitions of many technical terms, and still not be conscious of the attributes of good definitions, the general desirability of knowing the meanings of the words that one uses, or of the economy of thought that can be effected by inventing new concepts or improving existing terminology. He may be able to use simple algebra with facility, and yet be unaware of the role of symbolism in promoting economy and clarity of thought, ready manipulation of ideas, and increased ability to generalize. He can spend a whole year in the laboratory, and still be unable to render verbal the difference between observation and experiment. And if he does not verbalize and become functionally conscious of these various ideas, he will not put them to much use outside of the immediate situations in which they were learned.

Responsible citizenship.—As preparation for social leadership, the student should be helped toward the conviction that scientific attitudes and methods are applicable to all questions of fact, although with the expectation that the special technics which will be most effective will vary with the situation under investigation. He should see how physical science has evolved historically, so as to be able to predict probable courses of development in fields of knowledge and activity that are only beginning to have the status of a science. Especially valuable is a knowledge of science as a factor in history, and of the interplay of scientific and social forces, particularly those forces having profound effects on progress.

The student should be made acquainted with the specific character traits and attitudes exhibited by successful scientists in their scientific work[7] and with their group characteristics—prevalence of democratic procedures, free interchange of knowledge, cooperative effort on an international scale, and so on. If these qualities and practices are not simply moral virtues but indispensable scientific technics, then it may be that a greater premium should be placed on them in other, similar lines of endeavor where success comparable to that of the sciences is desired.

[5] Some of the contributions are mentioned in my paper on "Technical writing and editing," *Am. J. Physics* **13**, 99 (1945). For an extensive and illuminating semipopular treatment of the nature and structure of scientific language and its role in all modern education, see W. Johnson, *People in quandaries* (Harper, 1946).

[6] L. Bloomfield, *Linguistic aspects of science* (Univ. of Chicago Press), p. 1.

[7] See, for example, A. W. Hull, "Selection and training of students for industrial research," *Science* **101**, 157 (1945); comprehensive digest in *Am. J. Physics* **13**, 269 (1945).

To the superior student who plans to make economics, sociology, political economy or some related field his profession, physical science has a great deal to offer; witness, for instance, the current interest shown by social scientists in attempts of physicists to apply physical methods and ideas to social situations, as in the work of Bridgman[8] and of von Neumann.[9] But this offering will be most effective only if the teacher of these students has evident appreciation of the tremendous complexity and critical importance of social problems; of the fact that these problems still must be attacked by methods more closely akin to those of creative art than of the strict sciences; that, nevertheless, the logico-mathematical approach to social problems has also had some success; that if efforts in the latter direction often have been crude, they may be justified in situations so critical that workers in social fields cannot afford to await the development of procedures comparable in quality and power to those of the older, simpler sciences.

These workers are aware that long-range attacks on their problems also are necessary, and social students of today should prepare for them by gaining first-hand experience and critical understanding of the conditions and procedures that have made for progress in the older sciences. Conceivably, the methodology most suitable for social domains will ultimately turn out to differ fundamentally from existing scientific methods. This is by no means certain; but even if it should be true, the differences will not be clearly revealed and sharply defined until workers in these domains have given the existing methods appropriate and exhaustive tests.

Social students, especially, should find it profitable if in the physical science course emphasis were placed on subject matter illustrating ideas and generalizations such as the following:[10]

Those scientific advances that have led to applications of the greatest practical value have frequently been the result of work pursued without any regard for practical aims.

The most common phenomena are often the most difficult to explain, progress on them usually being made only when they are placed in contrast with related but uncommon phenomena. As A. N. Whitehead said, "It requires a very unusual mind to undertake the analysis of the obvious."

There are important distinctions between procedures that are merely laborious and those that involve inherent difficulties of a conceptual character; also between a complicated phenomenon that looks easy because knowledge of it is meager and a relatively simple phenomenon that appears difficult because the knowledge of it is detailed and highly developed.

To recognize a significant problem and to formulate it clearly is in general more difficult and requires greater creative ability than does the subsequent solution of the problem. Moreover, no problem can be adequately attacked until it has been clearly formulated.

Phenomena to which the principle of superposition is applicable or roughly applicable are the simplest to study and therefore have received the fullest treatment in physical science and the most emphasis in elementary courses. This has led to a rather widespread belief among laymen that only superposable phenomena are amenable to scientific treatment; hence, for instance, that the need for treating the whole rather than the parts in a social field constitutes a fundamental difference between social and all physical problems.

The question of the particular operations and materials to be used in carrying out any definite type of measurement is more fundamental than that of the particular system of units to be used in expressing the results of measurement.

Genuine theory, in the scientific sense, becomes possible only when there is an adequate background of carefully described and analyzed facts.

The choice of the particular concepts, definitions and principles upon which to base a theory, and of the particular logical order in which to develop it, is not unique. Several explanations developed from different points of view are valuable because they make for deeper understanding of a situation.

In developing a theory, the start should be made with modest or even trivial problems involving simple facts in a limited domain. The domain chosen should be one in which the results are already so well known that no theory actually is needed, for then the soundness of the theoretical approach can be readily checked.

In the early development of a theory, progress is often possible even though the concepts are not sharply defined or measurable with precision. The theory as it develops may point the way to refinements in these concepts or even to the invention of new concepts that are more fundamental and useful.

Once a theory has been made rigorous and general within its limited domain, extension to somewhat more complicated situations becomes possible, and there it may begin to reveal results that are not familiar. This calls for additional observations or experiments to check the new predictions, and thus begins the continual interplay be-

[8] P. W. Bridgman, *The intelligent individual and society* (Macmillan, 1938).

[9] J. von Neumann and O. Morgenstern, *Theory of games and economic behavior* (Princeton Univ. Press, 1944).

[10] The importance to social students of an understanding of a number of the items in this list is recognized by von Neumann and Morgenstern, reference 9, pp. 1–8. Several of these items appeared in the lists of common misconceptions about science prepared by P. Kirkpatrick and the author, *Am. J. Physics* **11**, 110, 163–4 (1943).

tween experiment and theory that is characteristic of any genuine science. The extensions into wider areas must of course be made with due regard for the greater dangers of extrapolation as compared with interpolation.

The use of mathematics in developing a theory not only promotes economy of effort and thought, but may be indispensable if the theory is to have greatest power. In this connection the question might be raised whether there is any conclusive evidence that the presence of human elements in social problems precludes the development of rigorous mathematical theory in this area.

The validity of any statement or generalization cannot be increased merely by re-expressing it in terms of mathematical symbols. Nor does such a mathematized statement, devoid of subsequent mathematical analysis, constitute a theory.

Mathematics is a powerful tool, and for this very reason must be handled skilfully and judiciously. If existing mathematical methods, after having been correctly applied to a new situation, prove to be inappropriate or difficult to handle, it may be found possible to modify the technics or even to develop a new and more suitable type of mathematics. Although there are promising mathematical methods already in existence that have not been exploited in the social fields, it is reasonable to suppose that ultimately there will also be need for the development of radically new mathematical methods for handling social theory.

Physics, the most fundamental and highly developed science, has as yet no all-inclusive theory serving to unify the science as a whole. Thus in fields immeasurably more complicated than physical science, to await the development of widely applicable theory is futile and precludes progress. In many social areas the first task apparently is still one of collecting enough data to make possible a start on real theory.

Esthetic appreciation.—In the realm of the creative arts, esthetic appreciation is a first step toward a deeper understanding of the role and meaning of art. In the sciences, the reverse process is the more usual one: increased awareness and understanding of phenomena and of man's physical and mental creations result in wider conceptions of what constitutes beauty—the beauty of countless physical phenomena that ordinarily pass unnoticed; the resplendency of a da Vinci or the more subdued beauty in the life and works of a Gibbs; the subtle beauty of a nice experiment, of a symmetrical equation, of an elegant mathematical proof, of a powerful generalization.

Of great practical value to the student is an understanding of the extent to which esthetic elements motivate and guide genuinely creative work of any kind; and of how such devices as esthetic balance effect economy of thought and make learning easier. Even beginners in physics unconsciously resort to such devices to enhance learning, as is evident in their predilection for physical equations expressed in the form of simple proportions and for mnemonic devices that are rhythmic in character.

Of indirect esthetic appeal is an appreciation of the extent to which physical advances can result in new tools and technics for the painter, the composer, the dramatist. Physical methods are also an integral part of most work in experimental esthetics, as in finding the relation between esthetic balance and mechanical systems in equilibrium, or in investigating esthetic judgments of color and of musical tone.

Much of this material related to esthetic appreciation can be introduced incidentally and with little expenditure of time. But the esthetic elements should be treated explicitly and with deliberate attempt to effect transfer to situations outside the sciences.

Personal adjustment; philosophy of life.—Of most obvious significance here is the convincing evidence which physical science can offer the student that it is possible to create working conditions in which the maintenance of certain high ideals of personal and social behavior is not only practicable but indispensable if workers are to be most productive and in all respects professionally acceptable. An effective approach to the discussion of this point is furnished by A. W. Hull's significant paper on the selection and training of students for industrial research.[11]

As many readers will remember, Hull lists four qualities sought in the hiring and retaining of research personnel, these being, in order of importance, *character*, *aptitude*, *attitude* and *knowledge*. As he says, if character traits such as self-discipline, courage, tolerance, honesty and generosity are first in importance, then they should be made a prime educational objective.

Second he places aptitude for research—curiosity, imagination, analytic ability—and here the function of the college is of course guidance. Third comes attitude—attitudes toward one's work, subject and fellow men; attitudes that will, among other things, make self-education possible. Finally comes knowledge, placed last not because

[11] Reference 7.

it is unimportant, but because of the high valuation given to the other three qualifications.

Hull makes the rather startling assertion that it is not so much the extra knowledge gained in graduate school as the changed attitude toward study that enables the man who has done graduate work to advance while those with only the undergraduate degree often reach a ceiling. One is reminded of the veteran undergraduate dean who, it is said, invariably concludes his annual report on the quality of the graduating class with the words: "Gentlemen, we've failed again; let's revise the curriculum." Another possibility, as Hull implies, is to revise the objectives of undergraduate work. If this were properly done, many curriculum problems might take care of themselves.

Incidentally, if the development and strengthening of character traits and attitudes are of first importance in preparing for a successful career as a scientist, then it follows that some of the prime objectives in the education of a physicist are among those also thought to be important in the education of the general student. This means, on the one hand, that the physicist who teaches general students does not have to go "out of his field" in order to emphasize these important objectives; and, on the other hand, that the teacher of physics majors who fails to emphasize these same objectives is badly missing the mark. Evidently the sort of course envisaged here as best for the nonscience student has features that should also be stressed even in advanced courses designed solely for physical science majors.

Examples can be cited of undergraduate professors who are rated by their students as poor classroom teachers, and yet have sent numerous young people on to successful careers in physics. Certainly the ability to impart knowledge is important, but evidently there are other qualifications that are essential for the most successful teaching.

Students, being thoroughly aware of the enormous successes of physical science, are likely to be impressed upon finding that the maintenance of high ideals of personal and group conduct is essential for scientific progress. This idea that a science provides a *natural* environment for the development and strengthening of desirable human qualities is important. Teaching character traits and attitudes is a ticklish matter, the outcome of which conceivably could be merely to help make prigs of the students—like the boy who set out to practise "every day and in every way getting honester and honester." The advantage of a science, in this respect, is that it provides, on the intellectual level, an atmosphere which in its healthy naturalness is strikingly similar to that of the playing field, where also the need for a strict code of conduct is generally recognized and accepted as a matter of course.

P. T. Orata,[12] in an illuminating article on the problem of transfer of training, makes effective use of the phrase "way of life" in describing school and college courses in which considerable transfer to extraschool situations is likely to occur. "School subjects will result in transfer to the social situations if, by proper instruction and organization, these subjects are made a 'way of life' and are so used by the student himself." This, Orata says, is what is meant by "humanizing education" in the concrete.

It is partly because of this need for making a subject a way of life that genuine laboratory work becomes almost indispensable in the physical science course. For the same reason a historical approach is likely to be effective, since, if properly employed, it gives the student opportunities to relive, even though vicariously, numerous significant events in the evolution of physical knowledge and thought; and, moreover, provides a natural way to acquaint him with the personal qualities, working habits and outlooks of those who have contributed most to scientific progress.

* * *

If little has been said about the amount of traditional subject matter that a student is likely to learn and retain in such a course, this is partly because there has been no extensive discussion here of such requisites of general education as "scientific understanding," which require explicit knowledge of subject matter. But even if such requisites were not emphasized, the amount needed cannot be appreciably reduced if the most is to be made of physical science in realizing other general objectives. Moreover, because of the stress that should be placed on the methodological and more theoretical aspects of physical science at the college level, the subject matter

[12] P. T. Orata, "Transfer of training and educational pseudo-science," *Math. Teacher* **28**, 265 (1935); *Ed. Admin. and Sup.* (Apr. 1935), p. 241; a digest adequate for the needs of most readers appears in *Am. J. Physics* **4**, 149 (1936).

chosen for presentation cannot possibly be superficial in character.

A teacher of students specializing in physics is not likely to present to them the several, major branches of the science as separate, unconnected disciplines. Similarly, it should be realized that, as Dean Cooper said, if general education really is to count, the general student also must gain a perspective and an awareness of the interrelationship of the courses he pursues—not only in physical science, but in the college as a whole. Only experience will show definitely how effective such a program will be; but it seems certain that it can be nothing less than an improvement over present efforts in schools and colleges to instruct people in the art of living in a modern society.

Preparation for living in a society in which science, in the broad sense, provides the basis for orientation, is of course a problem of far greater proportions than has been indicated here. Our discussion of personal adjustment, for instance, deals with only one small aspect of the general question.[13] That we will have a civilization of science is a foregone conclusion, assuming that we can achieve it ahead of disaster. This shift away from a prescientific culture may already be more appreciable than is commonly supposed. But the educational problems involved are still formidable, as seems evident merely from the observation that many scientists themselves find it difficult to utilize their training in dealing with their own personal and social problems. We have failed "to use it on us who use it." Education toward a civilization of science will have to be started earlier in life. But to get such an educational program for children into effect, there must be an informed adult opinion and teachers for the schools that are properly trained; so the start must be made in the college. If the task of completing the cultural shift before it is too late seems impossible, this can be regarded as still another reason why it should appeal to the best minds among physical scientists.

[13] See, for instance, W. Johnson, reference 5.

Origins and Ages of American Physicists

RAYMOND M. BELL
Washington and Jefferson College, Washington, Pennsylvania

IN recent months attention has been focused on physics and on American physicists. They form a relatively small group. It is interesting to examine their origins and ages. For this purpose all those listed in the seventh edition (1944) of *American Men of Science* as having *physics* for the major field have been tabulated. Somewhat similar surveys have been made by Birge[1] and Blackwood.[2]

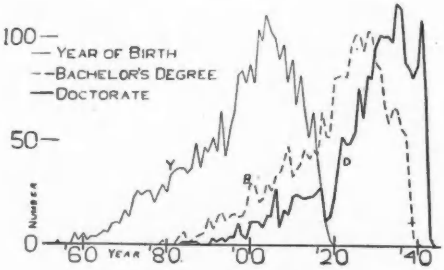


FIG. 1. Distribution of physicists.

In Fig. 1 are shown the number born in any one year, the number obtaining a bachelor's degree and the number obtaining the doctorate. To avoid confusion with honorary degrees, only Ph.D. degrees have been tabulated. The maximums of the three curves occur at 1904, 1928 and 1935, respectively. Since the curves are not symmetric, the median years are 1901, 1923 and 1931. This shows that the average physicist was 43 in 1944, 22 when he received his bachelor's degree, and 30 when he received the doctorate.

In Table I are listed the native states, and in Tables II and III the institutions granting the two degrees. Of the 2632 persons listed, 565 received only one of the degrees. There were 474 who received both bachelor's and doctor's degrees from the same institution. The birth dates are incomplete for 81 individuals.

[1] Birge, *Am. J. Physics* 9, 24 (1941).
[2] Blackwood, *Am. J. Physics* 12, 149 (1944).

As Table I shows, New York and Ohio have provided the largest number of physicists. Ontario comes first in Canada. A total of 131 are listed from Canada, 52 from Germany, 41 from Great Britain, 31 from the USSR, 93 from the rest of Europe, and 48 from other nations.

To determine the states contributing the highest ratio of physicists to population, the number per million persons has been calculated. The 1900 census was used as a basis; but, since some states increased greatly in population in the next ten years, ratios based on the 1910 census are also given for the leading states. The leading states are:

Over one million

	1900	*1910*
Ohio	49.3	43.1
California	49.2	30.7
Indiana	47.3	44.1
Iowa	41.7	41.8
Michigan	40.5	34.9
Wisconsin	39.6	35.2
Kansas	39.5	34.3
Massachusetts	37.1	30.9

Under one million

	1900	*1910*
Utah	119.2	88.5
Oregon	84.6	51.9
District of Columbia	71.7	60.4
Nevada	71.0	36.7
Colorado	70.5	47.6
Idaho	61.8	30.7
Montana	49.4	31.9
North Dakota	47.0	26.0
South Dakota	42.4	29.1

The New England states average 36.0. The national average is 33.5. On the basis of the 1900 census the leaders among the larger states are Ohio, California and Indiana; on the basis of the 1910 census the leaders are Indiana, Ohio and Iowa. Among the smaller states the leaders are Utah, Oregon and District of Columbia, using either census.

Of the 297 colleges granting 2401 bachelor's degrees, 59 have more than 10 individuals listed, 58 have 10 to 6, 53 have 5 to 2, and 127 only 1. Certain definite trends can be noticed before 1923 and after. In order to determine which colleges are outstanding in providing undergraduate training, the total number receiving bachelor's degrees in each college has been

TABLE I. Native states contributing 15 or more physicists; the numbers in parentheses are per million of 1900 population.

Before 1901		1901 and After	
Ohio	108	New York	127
New York	105	Ohio	97
Illinois	83	Pennsylvania	75
Indiana	79	Illinois	73
Pennsylvania	75	California	55
Massachusetts	64	Missouri	43
Iowa	60	Michigan	41
Michigan	57	Indiana	40
Wisconsin	46	Massachusetts	40
Missouri	35	Wisconsin	36
Kansas	33	Minnesota	34
Minnesota	25	Iowa	33
Virginia	24	New Jersey	26
Maryland	23	Oregon	26
New Jersey	23	Colorado	25
Maine	20	Kansas	25
Texas	19	Texas	25
California	18	Utah	21
Connecticut	18	Maryland	17
Nebraska	15	Virginia	17
		Nebraska	15

Totals

New York	232 (31.9)	Virginia	41 (22.1)
Ohio	205 (49.3)	Maryland	40 (33.7)
Illinois	156 (32.3)	Colorado	38 (70.5)
Pennsylvania	150 (23.8)	Oregon	35 (84.6)
Indiana	119 (47.3)	Utah	33 (119.2)
Massachusetts	104 (37.1)	Connecticut	32 (35.3)
Michigan	98 (40.5)	Nebraska	30 (25.7)
Iowa	93 (41.7)	Maine	27 (38.9)
Wisconsin	82 (39.6)	Kentucky	23 (10.7)
Missouri	78 (25.1)	North Carolina	23 (12.2)
California	73 (49.2)	West Virginia	21 (21.9)
Minnesota	59 (33.7)	District of Columbia	20 (71.7)
Kansas	58 (39.5)	South Dakota	17 (42.4)
New Jersey	49 (26.1)	Alabama	15 (8.2)
Texas	44 (14.4)	North Dakota	15 (47.0)

divided by the *total* enrolment in the institution as of 1923.[3] The leaders in institutions of various sizes are, in numbers per thousand:

Over 5000: Cornell, 9.5; Harvard, 7.2.
5000 to 2000: Massachusetts Institute, 23; Indiana, 14.
2000 to 1000: Oberlin, 24; Johns Hopkins, 20.
Under 1000: California Institute, 69; Reed, 57; Ripon, 49; Case, 36; Park, 29.

It is difficult to know the proper yardstick for comparing colleges granting bachelor's degrees. Another method is to divide the number by the total male enrolment in Arts and Sciences.[4] This has been done for the institutions listed below except for California Institute, Case and Massachusetts Institute, where the total regular en-

[3] *World almanac* (1924).
[4] *Biennial survey of education, 1922–24.*

TABLE II. Colleges granting 13 or more bachelor's degrees; the numbers in parentheses are per thousand of 1923 college enrolment.

Before 1923		1923 and After	
Indiana	39	Massachusetts Institute	41
Michigan	38	California Institute	31
Cornell	33	Wisconsin	27
Massachusetts Institute	33	Chicago	26
Harvard	24	Harvard	22
Johns Hopkins	24	Illinois	22
Toronto	24	Cornell	21
Oberlin	23	Michigan	21
Wisconsin	23	California	19
Yale	23	Oberlin	18
California	18	British Columbia	17
Northwestern	16	Columbia	17
Chicago	15	City College (N. Y.)	15
Dartmouth	14	Princeton	15
Columbia	13	Reed	15
Iowa	13	Case	14
Pennsylvania	13	Minnesota	14
		Ohio State	14
		Carnegie Institute	13
		Stanford	13

Totals

Massachusetts Institute	74 (23.3)	Texas	20 (4.0)
Michigan	59 (6.4)	British Columbia	19 (16.0)
Cornell	54 (9.5)	Colorado	19 (6.7)
Indiana	51 (13.7)	McGill	19 (6.8)
Wisconsin	50 (6.4)	Stanford	19 (6.3)
Harvard	46 (7.2)	Carnegie Institute	17 (4.0)
Chicago	41 (3.2)	Dartmouth	17 (8.2)
Oberlin	41 (24.5)	Reed	17 (56.7)
California	37 (2.4)	Virginia	17 (9.7)
Toronto	35 (7.3)	Washington (Seattle)	17 (2.9)
California Institute	34 (68.5)	DePauw	16 (11.4)
Illinois	32 (4.0)	Washington (St. L.)	16 (4.8)
Yale	32 (8.4)	Missouri	15 (2.3)
Johns Hopkins	31 (20.2)	Ohio Wesleyan	15 (7.9)
Columbia	30 (2.5)	Colorado College	14 (20.0)
Princeton	27 (11.4)	Miami (Ohio)	14 (9.6)
Minnesota	26 (2.9)	Wesleyan	14 (26.1)
Ohio State	26 (3.0)	Brigham Young	13 (7.2)
Case	23 (36.1)	Cincinnati	13 (3.1)
Pennsylvania	23 (3.1)	Kansas	13 (3.2)
City College (N. Y.)	22 (1.7)	Kentucky	13 (7.2)
Northwestern	22 (2.7)	Nebraska	13 (2.0)
Ripon	22 (48.9)	Park	13 (29.0)
Rochester	21 (27.9)	Pomona	13 (15.1)
Iowa	20 (2.7)		

rolment was used. Statistics are not available for Canadian institutions. The leaders are, in numbers per thousand:

Over 2000: Michigan, 17.2; Harvard, 15.7; Chicago, 14.2; Columbia, 14.2; Princeton, 12.7.

2000 to 1000: Massachusetts Institute, 60.6; Cornell, 42.3; Wisconsin, 31.6; Illinois, 28.9; Indiana, 26.9.

1000 to 400: Oberlin, 78.3; California Institute, 65.8; Johns Hopkins, 55.5; Case, 54.3; Rochester, 52.5.

TABLE III. Institutions granting 11 or more doctorates.

Before 1931		1931 and After	
Chicago	113	California	76
Cornell	94	California Institute	69
Johns Hopkins	94	Chicago	66
Columbia	50	Cornell	55
Harvard	47	Michigan	52
Michigan	45	Massachusetts Institute	50
California Institute	42	Wisconsin	46
Princeton	42	Harvard	44
Yale	42	Columbia	38
California	40	Yale	38
Wisconsin	37	Ohio State	37
Illinois	34	Johns Hopkins	36
Pennsylvania	28	New York University	36
Iowa	26	Illinois	32
Clark	18	Princeton	32
Göttingen	17	Indiana	29
Minnesota	15	Iowa	28
Indiana	14	Pennsylvania State	27
Ohio State	14	Virginia	21
Toronto	14	Minnesota	19
Berlin	13	Rochester	19
Purdue	13	Brown	18
New York Univ.	11	Pennsylvania	17
		Pittsburgh	15
		Washington (Seattle)	13
		Washington (St. L.)	13
		Stanford	12
		Toronto	12
		Iowa State	11
		Northwestern	11

Totals

Chicago	179	Pennsylvania State	28
Cornell	149	Toronto	26
Johns Hopkins	130	Stanford	22
California	116	Brown	20
California Institute	111	Pittsburgh	20
Michigan	97	Rochester	19
Harvard	91	Cambridge	18
Columbia	88	Clark	18
Wisconsin	83	Göttingen	18
Yale	80	McGill	18
Princeton	74	Northwestern	17
Illinois	66	Washington (St. L.)	16
Massachusetts Institute	59	Berlin	14
Iowa	54	Purdue	13
Ohio State	51	Washington (Seattle)	13
New York University	47	Iowa State	11
Pennsylvania	45	North Carolina	11
Indiana	43	Rice	11
Minnesota	34	Texas	11
Virginia	29		

Under 400: Reed, 107.5; Ripon, 90.0; Park, 74.8; Brigham Young, 68.5; Colorado College, 48.8.

The leaders in granting 2219 doctorates in physics have changed during the past few decades. California institutions are now ahead. Of the 105 universities, 39 have more than 10 individuals listed, 12 have 10 to 6, 18 have 5 to 2, and 36 have only 1. Of the last 36, only 15 are American universities. A total of 139 degrees are listed as having been granted by 37 European universities; 105 of these were granted before 1931.

Methods for Numbering and Cataloging Physics Equipment

Paul E. Martin
Wheaton College, Wheaton, Illinois

A PHYSICIST, if blessed with a good memory, may not need a record that indicates the locations of his equipment. But as the number of pieces multiplies, and the number of individuals who are responsible for the care and use of the instruments increases, one is usually driven to seek some method of reducing the time wasted in searching for misplaced pieces. Last year a questionary was sent to 210 physicists in colleges and universities throughout the country requesting information on: (i) the number system used; (ii) methods of affixing numbers to equipment; (iii) types of apparatus numbered; (iv) type of card file used. One hundred fifteen replies were received, 11 of them accompanied by letters describing in more detail the numbering and filing systems in certain departments.

Number Systems

Some departments have one system of numbers for laboratories, another for demonstration equipment, and still another for research apparatus. Certain large universities have numbers assigned to individual pieces of apparatus by the central inventory department or comptroller's office, where items are given consecutive numbers irrespective of the department receiving the apparatus. Brooklyn College uses the Kardex system. Holy Cross and Berea have tried modified Dewey Decimal systems. Various ingenious methods for identifying the general purpose of the apparatus are used, such as letters (say *L* for lecture, *R* for research) and colored dots or letters (say *brown* for research, *yellow* for photography). Muskingum, Oberlin, Rutgers and Wheaton are using hyphenated symbols that may include both letters and numbers. For example, in the designation *A-50-E-17*, *A* indicates the room, *50* the case, *E* the shelf or drawer and *17* the position number, or location of the apparatus on the shelf or in the drawer. Again, the first symbol may be the room number in the building or a departmental room symbol which may be stenciled on the inside of the door frame. Other symbols may be added to indicate the building or the floor of the building.

The writer has found it convenient to outline the position of each piece of apparatus on the shelf or in the drawer by means of a striping tool. Squares and rectangles only are used, and in the center of each is painted the position number of the equipment (Fig. 1). The shelf number is stenciled on the front edge of the shelf. The case number appears in a prominent position on the case, and the room number is stenciled on the inside of the door. Thus, as soon as a new instructor learns the sequence of the four symbols, he knows the departmental number for a missing piece of apparatus and can look up its number in the "check-out" book. Also from the number he knows where to go to put equipment in its proper place.

Affixing Numbers

There follows a summary of the replies regarding the methods of affixing numbers: painting, 70; engraving, 23; etching, 15; metal name plates, 13; stamped numbers (steel dies), 10; gummed labels and decalcomania, 15; no method, 21. The average is about one and a half ways of affixing numbers per institution, excluding the 21 institutions that reported no system used. One department "affixes the position of the equipment in the minds of the members of the staff." Paint is applied with an artist's brush, stencil, lettering sticks and rubber stamps. The colors most used are black or white, whichever gives the greater contrast with the equipment. Some use yellow on a black rectangular strip as background, or black on a white rectangular strip. The University of California finds India ink and artist's plain white paint ground in oil satisfactory.

Engraving of numbers is also common. A Corundum pencil[1] or diamond-tipped[1] scriber is

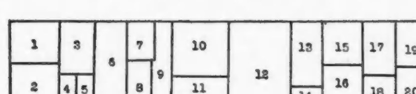


Fig. 1. Typical arrangement of items on shelf or in drawer.

[1] Central Scientific Co., Chicago 13, Ill.

used for hand engraving. A Burgess Vibro-tool[2] may become more popular for marking glass; the writer has tried it on the glass of vacuum tubes with apparent success.

Gummed labels, decalcomania and key tags are used by some institutions. Each of these may be purchased with printed material and the departmental numbers added. Large gummed labels, when attached to boxes containing all of the equipment for certain experiments, allow space for listing the various items. Decalcomania may be attached to a glass thermometer near the upper end without damaging the thermometer. Labels and decalcomania may be protected by covering them with shellac.

Several institutions use metal name plates.[3] These plates bear the name of the institution and the department, and in addition have sufficient space for the departmental number, which may be added by steel dies or by an addressograph (Oberlin) that embosses the numbers. Plates with borders in different colors are available. The spacing of the letters when steel dies are used may be made quite uniform by the use of a jig that holds the name plate in a fixed position relative to the die. The writer finds that a small metal shaper may be used for this purpose, the die fitting into the vertical groove moved horizontally by the means of a hand wheel, the number of revolutions of which determine the spacing of the letters.

One institution (University of California) employs steel dies ($\frac{1}{4}$-in. block letters) to mark wood and other nonbrittle substances, such as meter cases. The die is tapped lightly and swayed slowly, thus bringing part of the steel figure into contact at each blow. These indentations are then filled with an engraver's filler.[4]

Types of Apparatus Numbered

Many institutions do not attempt to number perishable or consumable articles such as glassware, vacuum tubes and small tools. Others do not number the equipment if the value is less than, say, \$2.00, \$3.00, \$5.00 or \$25.00. Certain small items, if similar, may be kept together in a box and the box given a departmental number. Aluminum pans, previously used as terminal box covers in airplanes, are convenient for this purpose; they may be purchased as government surplus. Slotted weights may be stored on the holder, which bears the departmental number. The position number may also be stamped on each of the individual items in a group. In case of the more delicate standard weights, no individual markings can be tolerated; therefore, care must be exercised to see that none of these are exchanged with those of another set.

Records

Most of the institutions having a number system for apparatus also have a card index file. The size of the card usually is either 3×5, 4×6 or 5×8 in. Although there is much variation in the printed material on the card, most of the cards have spaces for the name of the apparatus, departmental number, manufacturer's name and serial number, institutional number (if one is assigned), date of purchase and original cost. These cards should be light enough in weight to be typed easily. For convenience, one set of cards should be arranged alphabetically, and a duplicate set according to departmental numbers. Calibration sheets are kept in a separate file. The calibration sheets may be given consecutive numbers and recorded on the index card.

A systematic numbering and indexing scheme facilitates taking the inventory record periodically and locating borrowed equipment. The borrower should sign a card, or a "check-out" book, and record the number of the apparatus borrowed, the date, his department, and the estimated time of return. Some institutions place a duplicate card in a holder at the end of the case or lay it on the shelf from which apparatus was removed. The loan card should also bear the initials of the loaner. It is sometimes convenient to keep a permanent rather than temporary record of the loans, especially in the case of equipment that is likely to be damaged. When the apparatus is returned, one should take care to credit the loaner, perhaps using a credit slip.

* * *

While no one system of numbering and filing could be equally satisfactory for all departments, it seems desirable that some systematic plan be

[2] Burgess Battery Company, 184 N. Wabash, Chicago 1, Ill.

[3] Crowe Metal Name Plate Co., 3701 Ravenswood Ave., Chicago 13, Ill.; Etching Company of America, 1520 W. Montant Ave., Chicago, Ill.

[4] Binney and Smith Co., 41 East 42 St., New York, N. Y.

employed in each department. No system will work unless it is simple enough to be understood and consistently used by all who have the responsibility of removing equipment from storage and returning it again.

The writer wishes to thank the Editor of the *American Journal of Physics* for suggesting that a survey be made of practices in various institutions and to express his appreciation to all those who supplied information.

The Teaching of Units in Mechanics

C. F. Hagenow
Washington University, St. Louis 5, Missouri

DURING the period of army students' classes the methods of teaching the proper use of units, and their conversion, have again come under the scrutiny of textbook writers and teachers. There are, in particular, two aspects of the problem: (i) methods of solving problems in mechanics to insure the correct dimensions in the answer; (ii) methods of conversion of units within a given system or between two different systems of units; for example, between the cgs and the British engineering (Be) systems. One hears again the very old and familiar complaint that students never know "when to divide or multiply by g, the acceleration of free fall."

Now it has been my experience over many years that no such difficulty need ever arise if this subject is treated in a simple and logical manner. I proceed from the standpoint that one should teach only such systems of units in mechanics as are actually used in practice, namely, the systems used in the scientific and engineering worlds. Of the four systems of units—the absolute and gravitational within the metric and English systems—the only two of importance are the absolute cgs and the Be systems. It is true that in some laboratory experiments the gram is used as a unit of weight, but this need cause no confusion. In all these cases the precise unit employed is quite immaterial, since comparisons or ratios only are involved. No one cavils at the use of a centimeter of tube length as a unit of volume in a Boyle's law experiment.

In the Be system the fundamental units are force, length and time. The unit of force is the weight of the standard pound at (approximately) sea level and 45° latitude. Now the unit of mass is a *derived* unit in the Be system, obtained from the equation $m=w/g$. This relation also holds, of course, in the absolute cgs system. If the student learns Newton's second law he should know how to write the equation in any of the three forms, the form depending on the quantity for which he is solving.

Doubtless many engineers have never heard of the *slug*, the accepted term for the unit of mass in the Be system, and many of them know mass only as the ratio w/g. The fact remains, however, that in practice this unit of mass is used,[1] while the "poundal" is used nowhere except in some classrooms. Some time ago an English writer confessed his ignorance of just what "poundal" meant. He ventured the opinion that it might be some kind of yellow cake!

A criticism is often heard that the cgs and Be systems are not symmetric, since they differ in the choice of fundamental units. It is said that this situation is confusing and that students do not like it. But are we to sacrifice actual usage in the teaching of mechanics by introducing units used only in the solution of classroom problems merely because this procedure is found to be more palatable?

To come to the question of the solution of problems, the student's first question should be: which of these two systems of units am I going to use? In the Be system the answer is just this: the pound is always a unit of force; the mass of a given body is always w/g. I have met few in-

[1] For example, in the National Advisory Committee for Aeronautics, 25th Annual Report 1939, p. 242. Here mass density is expressed as "slugs per cu ft." In the 27th Annual Report, p. 90, moment of inertia is expressed in "slug-ft^2." In the 26th Annual Report, 1940, the table at the end of the volume gives the density of dry air at 760 mm and 15°C as ".002378 lb-ft^{-4} sec^2," which is dimensionally slugs per cu ft.

stances of the confusion that is said to be so prevalent. The student soon learns to recognize the fact that this unit of mass is not the legal unit but one necessitated by the engineer's choice of fundamental units.

In one textbook "lb" is used to denote a pound mass and "lbf" to denote a pound weight or force. Their respective counterparts in the metric system are "gm" and "gf." It seems to me to be a perfectly gratuitous complication to modify either "lb" or "g" by such additions. Even if such designations facilitated the solution of classroom problems, which I strongly deny, where else are they ever used? It is claimed that the use of "lb" and "lbf" is analogous to the use of "slug" as mass and "lb" as weight. But the symbols "lb" and "lbf" are, in my opinion, misleading. It seems as if a given body may at one time be a mass, at another, a force; a body has many attributes, but force is not one of its intrinsic properties. However, the strongest argument against the use of "lbf" is the fact that it is not employed by the very people who would seem to have the greatest need for it. They prefer w expressed in "lb" as their measure of force and w/g as their measure of mass.

Is it so difficult for the student to learn the meaning of a few well-defined units? If it is, I wonder how the addition of still more symbols can be of any help. It is important to point out to the student that each term must be of the same dimensions or, in simple language, *must be the same kind of thing*. It is unfortunately true that some students still try to add together cats and dogs; but is not that all the more reason why they should learn not to do it now?

The second topic, that of conversion of a given quantity into other units of the same or a different system, has had various treatments. A certain textbook has this example of a conversion of a mile into kilometers: 1 mi = 1 mi × (5280 ft/1 mi) × (12 in./1 ft) × (1 m/39.37 in.) × (1 km/1000 m) = 5280 × 12/(39.37 × 1000) km,[2] or 1.609 km. Here, again, there is the evident desire to wind up with the right dimensions, but what "pomp and circumstance." Now a system has long been in use that takes care of this necessity in a much simpler way. The foregoing example is almost too trivial to use, but here it is:

Let n be the number of kilometers in one mile. Since a *given length* must be the same in whatever units it is measured, we have

$$n \text{ km} = 1 \text{ mi, or } n = 1 \text{ mi/km} = 5280 \times 12/(1000 \times 39.37) = 1.609,$$

using the same conversion factors as before. [It would be simpler to use feet instead of inches from the table; thus

$$n = 1 \text{ mi/km} = 5280/(1000 \times 3.281) = 1.609.]$$

Consider the following example:

Convert 30 mi/hr to feet per second. In the textbook referred to, we have

$$30 \text{ mi/hr} = (30 \text{ mi/hr}) \times (5280 \text{ ft/1 mi}) \times (1 \text{ hr/60 min}) \times (1 \text{ min/60 sec}) = 30 \times 5280 \text{ ft/3600 sec} = 44 \text{ ft/sec.}$$ [3]

Now, more simply, let n be the number of feet per second; then

$$n \text{ ft/sec} = 30 \text{ mi/hr, and } n = (30 \text{ mi/ft}) \times (\text{sec/hr}) = (30 \times 5280/1) \times (1/3600) = 44,$$

the number of feet per second.

Finally, consider a more complicated case, in which we shall go back to the fundamental units:

Find the value of the gravitational constant in the Be system, given that its value in the absolute cgs system is 6.673×10^{-8} cm^3 g^{-1} sec^{-2}. Let K be the constant in the Be system; then

$$K \text{ ft}^3 \text{ slug}^{-1} \text{ sec}^{-2} = 6.673 \times 10^{-8} \text{ cm}^3 \text{ g}^{-1} \text{ sec}^{-2},$$

and so $K = 6.673 \times 10^{-8}$ (cm/ft)^3 (g/slug)$^{-1}$ sec^{-2}/sec^{-2}, or

$$K = 6.673 \times 10^{-8} \times \frac{1}{(30.48)^3} \times \frac{(453.6 \times 32.16)}{1} = 3.437 \times 10^{-8},$$

whose dimensions are, of course, those assumed for K to begin with.

All of the methods I have criticized are recommended on the ground of their being founded on good pedagogic principles. My only answer is the one already advanced, that the best pedagogy is one which conforms to actual practice and which employs the simplest means consonant with a clear understanding.

[2] On the right side of this equation the dimensions mile, feet, inches and meter cancel progressively, leaving only the kilometer.

[3] As in the previous example, like dimensions cancel on the right side of the equation, leaving only ft/sec.

Be tolerant. If you cannot or will not change your opinions, it is often possible to change the subject.—Anon.

The Physics of Radar (Continued)

SHERWOOD K. HAYNES* AND WILFRID J. JACKSON**
Radar School, Massachusetts Institute of Technology, Cambridge, Massachusetts

THIS is a continuation of an earlier article[1] in which the application to radar of some principles of physics is discussed. In the first article a block diagram of a typical search radar is given, and the basic principles of operation and the limitations of radar are discussed. A few typical circuits used in radar timing are described. Transmitting oscillators of the triode and the magnetron type are discussed, and the methods of pulse formation used in radar modulators are considered in some detail. The present article will discuss r-f components, antennas, receivers, indicators, sweep and range circuits, and direction indication as used in radar, with emphasis on the application of physical principles.

5. R-F SYSTEM

Two-wire unshielded transmission lines are rarely used in radar, for the radiation losses are prohibitive at all except the lowest frequencies. Therefore, either coaxial lines, in which the two conductors are concentric coaxial cylinders, or hollow pipes (wave guides) of circular or rectangular cross section are used. Low frequency alternating current theory cannot be applied to these lines because the wavelength of electromagnetic waves at radar frequencies is usually much less than the length of the lines. Thus the current cannot be considered constant throughout a two-wire line and its load.

A better approach is to consider that the transmission of energy consists of *current* and *voltage waves* traveling on the transmission line. Even this concept is not useful, however, in the case of hollow pipes, where there is only one conductor; here a useful concept is that of *electromagnetic waves guided by the pipe*. In fact, the guided electromagnetic wave concept can be applied also to coaxial and two-wire transmission lines. In *current and voltage wave theory* the instantaneous power transfer P across a plane perpendicular to the two conductors is given by

$$P = Vi, \tag{5-1}$$

where V is the instantaneous potential difference between the conductors at the plane, and i is the instantaneous current in one conductor at the plane. In *electromagnetic field theory* the flow of power across the plane is given by

$$P = \int_A \mathbf{S}\cdot d\mathbf{A} = \int_A (\mathbf{E}\times\mathbf{H})\cdot d\mathbf{A}, \tag{5-2}$$

where $\mathbf{S}$ is the Poynting vector, $\mathbf{E}$ is the electric field intensity, and $\mathbf{H}$ is the magnetic field intensity at the element of area $d\mathbf{A}$ of the plane. The correspondence which clearly exists between Eqs. (5-1) and (5-2) is discussed later in more detail.

Current and Voltage Wave Theory

Any two-conductor line has distributed inductance and resistance in the line and distributed capacitance across the line. Further, there will be leakage of charge from one conductor to the other, since the medium between them is not a perfect insulator. In Fig. 32 let the inductance per unit length of the transmission line be l, the resistance per unit length be r, the capacitance per unit length be c', and the shunt conductance per unit length be g. Applying Kirchhoff's laws to the path $ABB'A'$ and at the point A, one gets

$$r(dh)i + l(dh)\partial i/\partial t = dv, \tag{5-3}$$

$$c'(dh)(\partial v/\partial t) + g(dh)v = di. \tag{5-4}$$

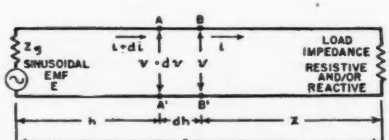


FIG. 32. Typical generator line and load.

* Now at Vanderbilt University, Nashville, Tennessee.
** On part-time leave of absence from Rutgers University, New Brunswick, New Jersey.
[1] S. K. Haynes and W. J. Jackson, *Am. J. Physics* **14**, 143 (1946).

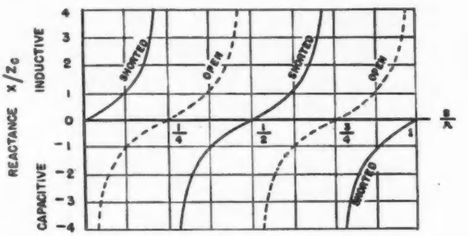


FIG. 33. Input reactance of short-circuits and open lossless lines *versus* electrical length.

Differentiating Eqs. (5-3) and (5-4) with respect to h and t, respectively, and combining the result with Eq. (5-4), one obtains

$$\frac{\partial^2 v}{\partial h^2}=rgv+(lg+rc')\frac{v}{t}+lc'\frac{\partial^2 v}{\partial t^2}, \tag{5-5}$$

which is the well-known equation for a damped wave. In radar the shunt conductance is negligible, and the lines are kept short so that the series resistance losses will be small. If terms containing r and g are neglected, Eq. (5-5) becomes

$$\partial^2 v/\partial h^2 = lc'\partial^2 v/\partial t^2, \tag{5-6}$$

which is the familiar form of the one-dimensional wave equation. A corresponding equation for i results when v is eliminated between Eqs. (5-3) and (5-4):

$$\partial^2 i/\partial h^2 = lc'\partial^2 i/\partial t^2. \tag{5-7}$$

If Eqs. (5-6) and (5-7) are solved for v and i, the following interesting results are obtained:

(1) Both current and voltage waves travel with a velocity

$$v=(lc')^{-\frac{1}{2}}=c(\mu_r\epsilon_r)^{-\frac{1}{2}}, \tag{5-8}$$

where c is the speed of light in free space and μ_r and ϵ_r are the relative permeability and dielectric constant of the medium between the conductors.

(2) In general there will be current and voltage waves in both directions, for reflection will occur at every point of discontinuity. If there is no reflected wave and the line is lossless, absorption of power by the load will be complete and, since there is no damping, the maximum values of voltage and current will be the same at all points on the line, while the phase of the oscillation will be delayed 360° for each wavelength toward the load. On the other hand, if total reflection occurs, so that the amplitudes of the incident and reflected waves are equal, then no power is absorbed by the load, and a *standing wave pattern* results with nodes and loops of both current and voltage. The current nodes are always separated from the voltage nodes by a quarter wavelength.

In general, some reflection occurs, but the amplitude of the reflected wave is smaller than that of the incident wave. Then there are minimums instead of nodes, and the loops are smaller. In such situations a useful quantity is the *standing wave ratio*, which can be defined as the ratio of the maximum voltage or current to the corresponding minimum. For efficient power transmission over a long line, a standing wave ratio close to unity is necessary.

(3) If a line were infinitely long, there would be no reflected wave. But even a line of finite length can be terminated in such an impedance that no reflection will occur. This impedance is denoted by the symbol Z_c and is called the *characteristic impedance* of the line. From the solutions of Eqs. (5-6) and (5-7) it can be shown that, for a lossless line,

$$Z_c=(l/c')^{\frac{1}{2}}. \tag{5-9}$$

(4) With a sinoidal emf applied, the current and voltage at all points vary sinoidally with the frequency of the emf. The only difference in the behavior of either the current or the voltage at different points of the line, once the *steady state* is reached, is in their amplitude and their phase. Most treatments of lines use the rms (effective) value of the amplitude and employ complex notation to indicate phase. Then the ratio of the complex effective voltage to the complex effective current at a point a distance x from the load can be shown to be[2]

$$\bar{V}/\bar{\imath}=Z_c[Z_R+jZ_c\tan(2\pi x/\lambda)]/[Z_c+jZ_R\tan(2\pi x/\lambda)]. \tag{5-10}$$

(5) Two important ways to terminate a line, other than by the characteristic impedance, are an open circuit and a short circuit. From Eq. (5-10) the input reactance of lossless lines with such terminations can be found as a function of the electrical length, s/λ, where s is the length of line. The results are shown in Fig. 33. Shorted

[2] See Brainerd, Koehler, Reich and Woodruff, *Ultra high frequency techniques* (Van Nostrand), pp. 343–346. Equation (5-10) is obtained by dividing Eq. (11-9) by Eq. (11-10) and setting γ equal to $2\pi/\lambda$.

lines are particularly important as resonant elements in triode transmitters (SEC. 4), as impedance matching elements and as supporting stubs for coaxial lines. Both shorted and open lines are important in duplexers and rotating joints (Fig. 34).

(6) Equation (5-10) is usually plotted either as a so-called *circle diagram*[3] or as a Smith chart,[4] and most problems are solved graphically by using these charts.

A possible coaxial radar transmission line system is shown schematically in Fig. 34. On the left is shown the portion of the line next to the transmitter-receiver; on the right, that next to the antenna. The length of line connecting the two portions is generally quite long. In order that the losses on this long connecting section be small, the standing wave ratio for this section should be as close to unity as possible. If the impedance of the antenna is not in itself the characteristic impedance of the line, reflection can be prevented by the use of tuning stubs. Since a lossless stub is always reactive (Fig. 33), each stub adds a pure susceptance to the admittance of the elements to its right. The right-hand stub adds sufficient susceptance to the antenna admittance so that, *at the left-hand stub*, the combined conductance is the characteristic conductance, $1/Z_c$, of the transmission line. The left-hand stub then adds sufficient susceptance to neutralize the residual susceptance of the elements to its right. Details of this and other matching problems will be found in more complete treatments.[3-5]

Double stubs may also be used at the transmitter end of the r-f system to insure the proper loading of the transmitter. "Proper loading" in this case means maximum power output consistent with satisfactory tube life and good frequency stability.

Adjustable double stubs are not always used for impedance matching. However, the problems of proper impedance relations at the antenna and at the transmitter must be solved for each radar either by some fixed prematched arrangement or by adjustable elements on the system.

The center conductor of coaxial lines may be supported by solid dielectric, by porcelain beads or by quarter-wave stubs as shown in Fig. 34. Solid dielectrics are not sensitive to frequency changes, but cause considerable loss. Beads break down if wet or dirty and introduce reflections and are somewhat fragile. Quarter-wave stubs are most useful for short wavelengths, since only then will they be of practicable size.

A rotating joint enables the r-f energy to enter the rotating r-f system without appreciable loss or reflection. Physical contact of sliding surfaces is impracticable owing to wear and to poor contact. The arrangement shown in Fig. 34 connects the center conductors by a coaxial line within a line. The approximate open circuit at the inner end of the line is transformed into a radiofrequency short circuit a quarter-wave away (Fig. 33) at the outer end. In a similar manner the set of three coaxial lines outside the outer conductor insures a radiofrequency short at the junction of the outer conductors.

The "bazooka" at the antenna (Fig. 34) is another illustration of the usefulness of quarter-wave short-circuited lines. The antenna in Fig. 34 will tend to radiate guided waves along the outside surface of the coaxial line. The quarter-wave "bazooka" reflects this energy with such a phase as to present a high impedance to the antenna. Hence, in effect, it uncouples the outside of the line from the antenna.

The duplexer connects the receiver to the antenna at all times except during the transmitted pulse. The duplexer is a quarter-wave coaxial line shorted at one end and having a spark gap (open circuit) at the other end. The gap is

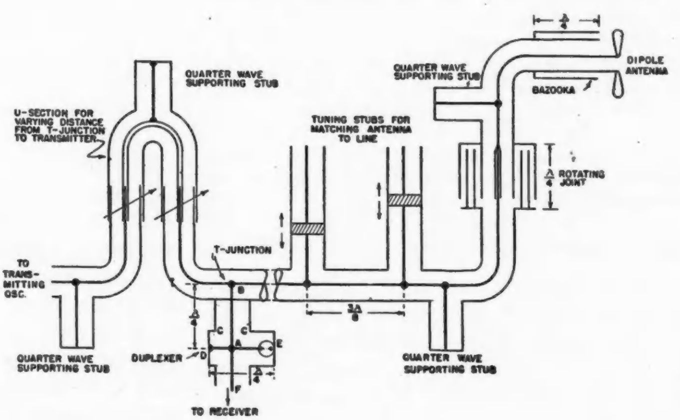


FIG. 34. A possible coaxial radar transmission line system.

[3] Wing and Eisenstein, *J. App. Physics* **15**, 615 (1944).
[4] P. H. Smith, *Electronics* **17**, 130 (1944).
[5] Slater, *Microwave transmission* (McGraw-Hill), pp. 42–78.

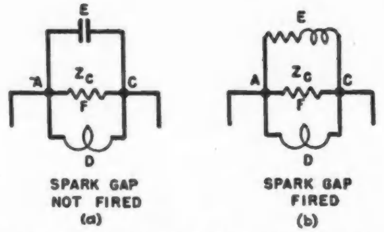


FIG. 35. Equivalent circuit of the duplexer, Fig. 34.

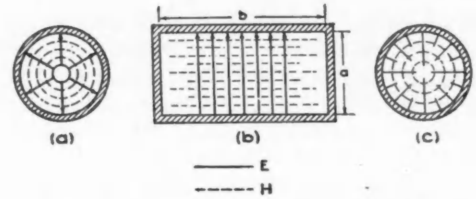


FIG. 36. Cross sections of r-f transmission lines.

usually enclosed in a tube of gas at reduced pressure to decrease the breakdown and operating voltage. The equivalent circuit of the duplexer of Fig. 34 is shown (gap not fired) in Fig. 35(a). The section ADC acts as a very small inductance, whereas the section AEC acts as a large capacitance. Since the total length of the line from D to E (Fig. 34) is one-quarter wavelength, it is seen from Fig. 33 that the reactances of these sections are equal in magnitude and opposite in sign. Hence the impedance between A and C is just the input impedance of the receiver, which is usually equal to the characteristic impedance of the line. With the gap fired [Fig. 35(b)], the right-hand end of the line is no longer capacitive and no longer neutralizes the very low inductive reactance of the section ADC. Thus, when the gap is fired, the input impedance between A and C is nearly zero.

When the transmitter oscillates, the gap breaks down and the impedance between A and C, being nearly zero, prevents application of a large voltage to the receiver. From the T-junction B, one-quarter wavelength away, the duplexer looks like an open circuit and very little transmitted energy is diverted toward the receiver.

When a signal is received by the antenna, it will travel easily through the duplexer to the receiver, for the input impedance is Z_c. In order that the transmitting oscillator neither absorb received energy nor cause received energy to be reflected from the T-junction to the antenna, the impedance looking toward the transmitter from the T-junction must be high. In most cases the transmitter will reflect received signals in such a way as to give a high standing wave ratio on the line between the transmitter and the T-junction. By adjusting the U-section (Fig. 34), the length of line can be varied, so that there will be a voltage maximum, and hence an impedance maximum, at the T-junction.

Electromagnetic Field Theory of Guided Waves

From Maxwell's equations, for a medium containing no volume concentrations of charge, the following well-known differential equations for the behavior of the electric field intensity $\mathbf{E}$ and the magnetic field intensity $\mathbf{H}$ can be derived:

$$\nabla^2\mathbf{E}=\sigma\mu\frac{\partial\mathbf{E}}{\partial t}+\epsilon\mu\frac{\partial^2\mathbf{E}}{\partial t^2}, \tag{5-11}$$

$$\nabla^2\mathbf{H}=\sigma\mu\frac{\partial\mathbf{H}}{\partial t}+\epsilon\mu\frac{\partial^2\mathbf{H}}{\partial t^2}. \tag{5-12}$$

Here μ is the permeability of the medium, ϵ is it dielectric constant, and σ is its conductivity; mks rationalized units are used. The solution of these equations predicts the existence of electromagnetic waves.

The transmission of waves through conductors having any of the three cross sections shown in Fig. 36 can be described by solving Eqs. (5-11) and (5-12) both for the enclosed dielectric and for the conducting medium, and applying the following boundary conditions:

(a) The tangential component of $\mathbf{E}$ is the same on both sides of the boundary.

(b) The normal component of $\mathbf{B}[=\mu\mathbf{H}]$ is the same on both sides of the boundary.

(c) The tangential component of $\mathbf{H}$ is continuous on both sides of the boundary.

(d) The discontinuity in the normal component of $\mathbf{D}[=\epsilon\mathbf{E}]$ equals the surface charge density.

Certain general and specific results are obtained from the solutions of these equations.

General Results

(1) The electromagnetic field in the metallic boundaries falls off to 0.37 of its surface value at a skin depth δ given by

$$\delta=\sqrt{(1/\pi\sigma\mu f)}, \tag{5-13}$$

where f is the frequency of the radiation. For copper and other good conductors, if f is about 10^2 to 10^4 megacyc/sec, then δ is of the order of 10^{-6} m. For most purposes, the walls may be considered to be perfectly conducting and the skin depth δ, to be zero. Then the boundary conditions become:

(a) The tangential component of $\mathbf{E}$ is zero on both sides of the boundary.

(*b*) The normal component of **B** is zero on both sides of the boundary.

(*c*) The tangential component of **H** in the dielectric equals the surface current density in the conductor.

(*d*) The normal component of $\mathbf{D}[=\epsilon\mathbf{E}]$ in the dielectric equals the surface charge density on the conductor.

(2) When Eqs. (5-11) and (5-12) are solved for the conductors of Fig. 36, many different possible solutions are obtained. These solutions yield various field configurations, called *modes*, that vary over the cross section of the dielectric and also along the axis of the guide. For the wave guides of Fig. 36, (*b*) and (*c*), if the frequency is sufficiently high the field configurations are propagated along the axis of the guide as waves, but for lower frequencies the field intensities are rapidly attenuated. The simple modes illustrated in Fig. 36 are of the propagated type and are usually the desired modes.

For the wave guides of Fig. 36, (*b*) and (*c*), there is, for a given frequency and mode, a critical, linear dimension of the wave guide below which the mode will not be propagated. Fortunately, the desired modes (*b*) and (*c*) will pass through a smaller guide than almost all other modes. To obtain propagation in the proper mode, therefore, the guide is made large enough to pass the desired mode but small enough to exclude as many other modes as possible. The method of excitation is so chosen as to eliminate other undesired modes.

The critical dimension of a hollow-pipe wave guide is of the same order of magnitude as the free-space half-wavelength. Hence, hollow pipe guides are impractical for wavelengths greater than a few centimeters. Since the mode shown in Fig. 36(*a*) for the coaxial line is transmitted at all frequencies, coaxial lines can be used for longer wavelengths and still be of reasonable size.

Whenever practicable, a guide is used in preference to a coaxial line. It is simpler to construct, since it consists of only one conductor. The copper loss of a hollow pipe is less than that of a coaxial line because of the larger skin effect in the coaxial line owing to the small surface area of the inner conductor. The breakdown voltage of a wave guide obviously is higher than that of a coaxial line of similar dimensions; the distance across the largest single-mode pipe is much greater than the distance between the inner and outer conductor of the largest single-mode coaxial line.

(3) As already indicated, for a given mode there is a relation between the frequency (or free-space wavelength) and the smallest guide in which that mode will be propagated without excessive attenuation. Thus for a given guide there is a critical free-space wavelength λ_c for each mode which cannot be exceeded if the mode is to be propagated.[6]

In contrast to the fundamental mode in a coaxial line, which is propagated with free-space velocity if the dielectric is air, the hollow pipe modes are propagated with phase velocities that vary with frequency and that always exceed the free-space velocity. The corresponding wavelengths, called *guide wavelengths* λ_g, are therefore longer than free-space wavelengths λ_a. For every mode it can be shown that

$$\lambda_g = \lambda_a/[1-(\lambda_a/\lambda_c)^2]^{\frac{1}{2}}. \tag{5-14}$$

These results can be easily understood for the fundamental mode of a rectangular guide by reference to Fig. 37, which shows two sets of

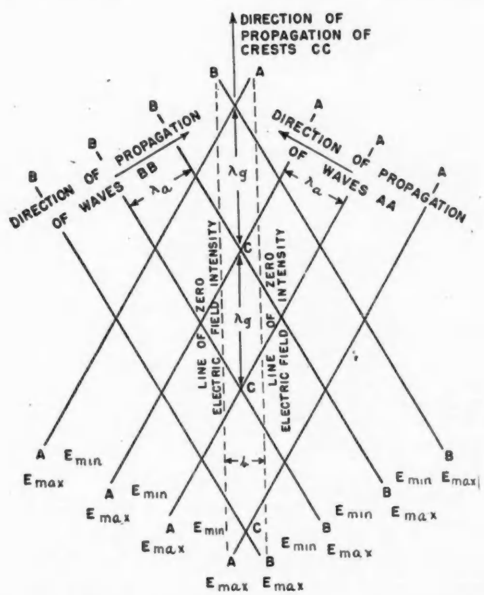


FIG. 37. Two sets of plane wave fronts crossing obliquely.

[6] For detailed descriptions of the more important modes and dimensions of the wave guide, see, for example, reference 5, chap. 6.

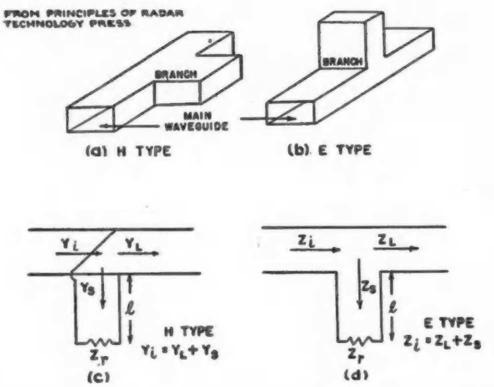


FIG. 38. T-junctions and their transmission-line equivalents.

plane wave fronts (loci of maximum electric field intensity) crossing obliquely. While the wave fronts move with a velocity v_a and have a wavelength λ_a, their intersections C move with a velocity $v_g[>v_a]$ and have a wavelength $\lambda_g[>\lambda_a]$. The dashed lines indicate loci of zero field intensity in the wave pattern. If perfectly conducting walls are placed along the dashed lines, the field between the lines will be unaltered. Calling the distance between the dashed lines b, one can show that

$$\lambda_g = \lambda_a / [1-(\lambda_a/2b)^2]^{\frac{1}{2}}; \qquad (5\text{-}15)$$

and from Eq. (5-14) it follows that, for this mode,

$$\lambda_c = 2b. \qquad (5\text{-}16)$$

The interpretation for other modes is similar but more complicated.

Since, by Eq. (5-15), the guide wavelength and phase velocity vary with the free-space wavelength, not all the component frequencies of a pulse of r-f energy will have the same phase velocity. The pulse will move with a group velocity given by

$$v_g = c^2/v_p = (\lambda_a/\lambda_g)c, \qquad (5\text{-}17)$$

where v_p is the average phase velocity of the components of the pulse, and c is the speed of light.

(4) Many of the ideas discussed in connection with transmission lines* are also at least qualitatively applicable to wave guides if the electric and magnetic field strengths are considered analogous, respectively, to potential difference and current in a coaxial line, as was suggested in connection with Eqs. (5-1) and (5-2). Thus, in the light of this analogy, the impedance at a point in the line varies with the ratio of **E** to **H**, or

$$Z = kE/H. \qquad (5\text{-}18)$$

Although certain difficulties and discrepancies arise in determining the proportionality constant k in Eq. (5-18), relative impedances along a wave-guide system frequently can be found by use of this concept of impedance and by use of either Eq. (5-10), the circle diagram or the Smith chart. The wavelength used in these calculations must be the guide wavelength λ_g.

The qualitative analogy to transmission lines for the rectangular guide and mode of Fig. 36(b) can be extended by considering the center portion of the top and bottom conductors of the guide to be analogous to the conductors of a two-wire transmission line. These conductors may be considered roughly to be supported continuously throughout their length on each side by quarter-wave stubs. The width of the center portion which comprises the analogous transmission line is $b-\frac{1}{2}\lambda_a$.

(5) Branch connections from a rectangular guide, Fig. 36(b), can be of the two types shown in Figs. 38(a) and (b). The equivalent transmission lines are shown in Fig. 38(c) and (d). It follows from the qualitative analogy of the preceding paragraph that the H-type should be equivalent to a parallel connection and the E-type to a series connection. It should be noted that the junction points and admittance Y_s in Fig. 38(c) correspond to the center of the main wave guide of Fig. 38(a), and not to the aperture in the side wall which is effectively a quarter wavelength away.

A possible wave-guide r-f system is shown schematically in Fig. 39. The r-f energy from the magnetron reaches the rectangular wave guide through a short section of coaxial line. The line is coupled to the guide by a special antenna so arranged as to introduce negligible reflection. Both the radiation resistance and the reactance of the antenna depend on the depth of penetration of the antenna into the guide and on the distance from the antenna to the closed (left-hand) end of the wave guide. The proper distances (approximately $\lambda_g/6$ for each) are usually de-

* Resulting from the solutions of Eqs. (5-6) and (5-7).

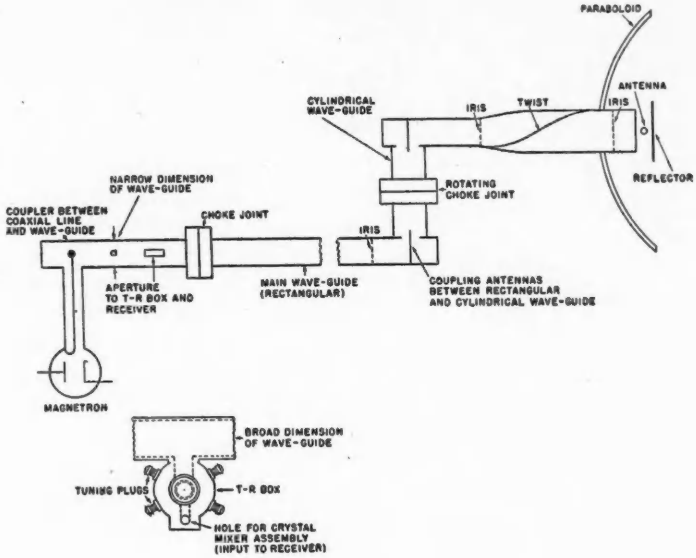


FIG. 39. A possible wave guide r-f system.

termined in the laboratory and built permanently into the radar.

The r-f energy then passes to the right along the rectangular wave guide whose narrow dimension is shown in Fig. 39. As the first of this energy reaches the aperture leading to the *T-R* box, the firing of the spark gap in the *T-R* box effectively closes the aperture in a manner similar to the action of the duplexer in Fig. 34.

The choke joint (Fig. 40) passes r-f energy without mechanical connection from one section of wave guide to another. Such a joint is useful in keeping outside vibrations from reaching the transmitter and receiver and in making possible the removal of sections of the wave guide for repairs. Since the flow of energy is concentrated near the center of the wave guide, the choke joint must prevent energy flow in the direction ABC. In this direction the choke forms an E-type junction with the main wave guide. The circular slot forms a shorted coaxial line which makes an E-type junction with the interflange region at B. By making the input impedance to the slot infinite and the distance AB equal to $\lambda_g/4$, a short circuit will appear at A, as is desired. [$Z_s=0$ in Fig. 38(d).]

From the choke joint the rectangular wave guide goes to the antenna assembly—more than 100 ft in some cases. As in the coaxial r-f system, a rotating joint is necessary (Fig. 39); hence a circular rather than a rectangular guide must be used. The energy is transferred from the rectangular guide of Fig. 36(b) to the cylindrical guide of Fig. 36(c) by means of an antenna that is parallel to the electric field in both guides. Elimination of reflections is accomplished by adjusting the lengths of the antenna and of the rectangular guide to the right of it (Fig. 39) and, if necessary, by introducing an *iris* (to be discussed later) to the left of the antenna. The rotating joint itself is merely a rotating choke joint for a cylindrical wave guide. After passing through the rotating choke joint the r-f energy returns to a rectangular guide through another coupling antenna and iris.

Although wave-guide stubs analogous to coaxial stubs could be used for impedance matching, the perforated diaphragms, or irises, shown in Fig. 41 are preferable since they are less bulky. An iris, depending on its orientation, will act as either a capacitance or an inductance shunted across the equivalent transmission line representing the main wave guide. In Fig. 39 irises are shown not only as matching aids at the entrance and exit of the rotating joint, but also at the input to the antenna so that the latter will receive energy from the guide without reflection.

After leaving the rotating joint the wave guide may be twisted (Fig. 39), if it is desired that the polarization from the antenna be horizontal. The antenna shown in Fig. 39 is a half-wave dipole normal to the page with a reflector in front of it. The antenna and small reflector scatter the energy from the wave guide in such a way as to illuminate the parabolic reflector nearly uniformly.

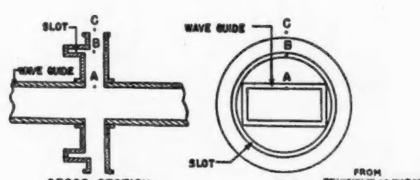


FIG. 40. Choke joint.

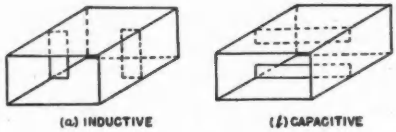


FIG. 41. Wave-guide irises.

The path of any received energy and of the transmitted energy are the same between the *T-R* box aperture and the antenna. At the aperture the received energy would divide except for the fact that the magnetron, when not transmitting, terminates the transmitter end of the line with a reactive impedance that sets up a high standing-wave ratio in the line between magnetron and *T-R* aperture. The *T-R* aperture is located at such a distance from the magnetron that the impedance at the aperture and looking toward the magnetron is very high. Thus, if the *T-R* aperture is properly matched to the wave guide, virtually all the received energy will pass to the receiver.

Microwave *T-R* boxes differ from *u-h-f* duplexers in the following respects:

(1) A resonant cavity (Fig. 42; Fig. 39, top view) is used instead of a tuned transmission line. This cavity oscillates with circular magnetic flux around the spark gap and with approximately vertical electric field. The magnetic field is strong at the outside and weak at the center, and the electric field is very strong in the gap and very weak outside. The gap, therefore, is at a very high impedance level (E/H large) compared to the input aperture from the wave guide or to the mixer assembly of the receiver. The operation and equivalent circuits of the cavity are similar to those of the duplexer of Fig. 34, where the gap is at a high voltage and impedance level compared to the input and output. Thus, during the transmitted pulse the gap is fired, and a very low impedance exists across the aperture between the wave guide and *T-R* box, and very little energy will reach the receiver. When signals are being received, the gap is not fired, the cavity resonates, and the input impedance to the *T-R* box, which must roughly match the wave-guide impedance, is approximately equal to the input impedance of the receiver.

(2) Crystal mixers require better protection against the transmitted energy that passes the *T-R* box before the gap breaks down and also against the energy that leaks through after the gap fires. Proper gas pressure in the tube containing the gap and use of a negative "keep alive" electrode (Fig. 42) to furnish a few ions, makes rapid breakdown and low operating voltage certain.

(3) The *T-R* box must de-ionize rapidly after the transmitted pulse has passed, and must not attenuate the received signal excessively. A combination of hydrogen and water vapor in the tubes gives the best over-all results in these respects. The high rate of volume recombination of the water vapor ions aids greatly in rapid recovery.

A combination of gases (hydrogen or argon) with water vapor in the tubes gives the best over-all results in these respects. The high rate of electron capture by water vapor molecules aids greatly in rapid recovery.*

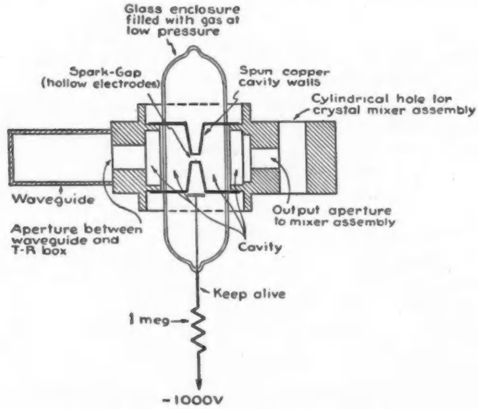


FIG. 42. Microwave *T-R* box, side view.

6. ANTENNAS AND PROPAGATION

The purpose of a radar transmitting antenna is to direct the energy from the r-f transmission system into space as a narrow beam of electromagnetic energy. The transition from guided electromagnetic waves to free-space waves is usually either through one or more half-wave antennas (known as dipoles) such as those of Figs. 34 and 39, or through the open end of a wave guide (Fig. 43), which is a type of electromagnetic horn. Since neither a single dipole nor an open wave guide is highly directive, the energy is concentrated into a narrow beam by use of an

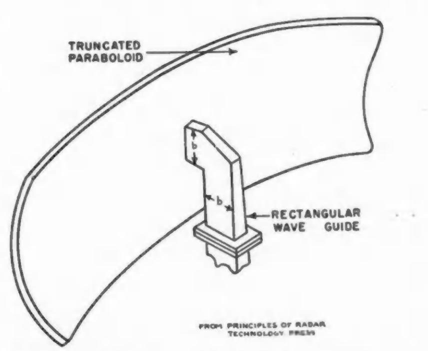


FIG. 43. Rectangular wave-guide horn feeding a truncated parabolic reflector.

* Margenau, McMillan, Dearnley, Pearsall, and Montgomery, *Physical Rev.* **70**, 349–357 (1946).

array of dipoles, backed by a plane reflector, or by placing the horn or one or more dipoles at the focus of a parabolic reflector.

The exact calculation of the radiation pattern of a metallic center-fed half-wave antenna by means of Maxwell's equations and the first set of boundary conditions of SEC. 5 encounters prohibitive difficulties.[7] A simpler and fairly accurate method is to assume a sinoidal current distribution along the antenna and to compute the resulting radiation pattern (Fig. 44). The radiation pattern from an open rectangular wave guide is shown in Fig. 45.

Directive mechanisms.—The use of a linear array of dipoles to obtain a narrow beam is similar to the use of a diffraction grating[*] to obtain narrow spectrum lines. In radar, however, the zero order is used; the higher orders are unwanted inasmuch as only *one* narrow beam is desired. If m identical sources are spaced a distance d apart (Fig. 46), the intensity of the radiation in a direction α is equal to that from a single source multiplied by a factor F^2:

$$F^2(\alpha) = \sin^2\left(\frac{m\pi d \cos\alpha}{\lambda}\right) \Big/ \sin^2\left(\frac{\pi d \cos\alpha}{\lambda}\right), \quad (6\text{-}1)$$

which is a familiar factor in grating theory.[8] The factor F^2 has a maximum value equal to m^2 for each value of α for which the denominator goes to zero and has smaller maximums in the neighborhood of values of α for which the numerator is unity. The elimination of all principal maximums

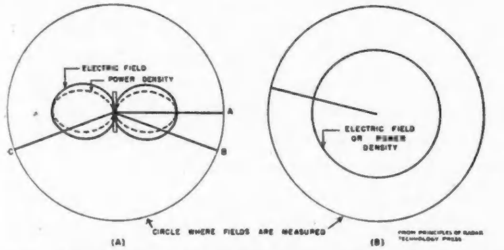


FIG. 44. Pattern for a half-wave antenna: (*A*) in the plane of the antenna; (*B*) in the plane perpendicular to the antenna.

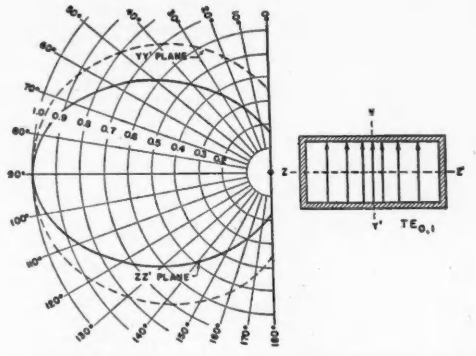


FIG. 45. Radiation pattern of a rectangular wave guide.

except that for $\cos\alpha = 0$ is accomplished by making the ratio d/λ less than unity. Usually the sources are spaced one-half wavelength apart ($d/\lambda = \frac{1}{2}$). When a narrow beam, both horizontally and vertically, is desired, dipoles which radiate in phase are arranged in both horizontal and vertical rows and the pattern of a single dipole must be multiplied by two factors, one for each direction. The usual number of dipoles in a row varies from four to eight or more.

Such a two-dimensional array will have radiation maximums normal to the plane of the array, both forward and backward. A plane wire screen placed approximately one-quarter wavelength behind the two-dimensional array not only appreciably reduces back radiation but adds to the directivity of the pattern.

As far as forward radiation is concerned, the action of the screen can be thought of as replaced by an array of image dipoles situated one-quarter wavelength behind the *screen*. In order that the radiation electric field intensity may be zero at the screen, the radiation from the image dipoles must be 180° out of phase with that from the actual dipoles. In this simple case it is obvious that the field of image dipoles one-half wavelength behind actual dipoles and leading the latter by 180° will be in phase with the field of the

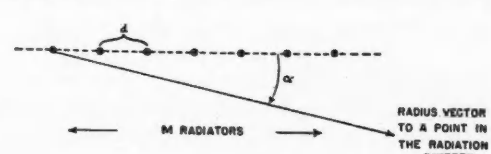


FIG. 46. Space pattern, showing angle α.

[7] Reference 5, pp. 209–232.
[*] The grating analogy is accurate only if the pattern of any one dipole is not altered by the presence of the others. Interactions between dipoles will be neglected in this treatment.
[8] Drude, *The theory of optics* (Longmans Green), p. 222, Eq. (85).

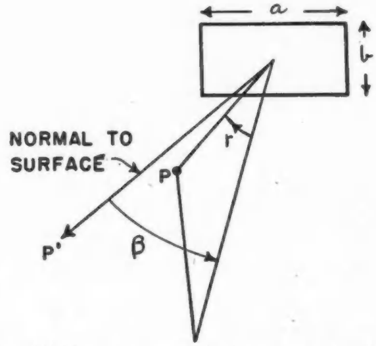


FIG. 47. Space pattern, showing angles β and γ.

actual dipoles in the forward direction. However, the whole pattern is more clearly revealed by a factor analogous to Eq. (6-1) that takes into account the difference in phase of identical sources which, in this case, are the object and image two-dimensional arrays. If α in Fig. 46 is measured from the lagging end of a line of identical sources whose phase difference is φ cycles, the power density of the radiation pattern of a single source is multiplied by[9]

$$G^2(\alpha)=\sin^2 m[\pi\varphi-\pi(d/\lambda)\cos\alpha]/\sin^2[\pi\varphi-\pi(d/\lambda)\cos\alpha]. \quad (6\text{-}2)$$

In the case of the reflector one-quarter wavelength behind the dipoles, $\varphi=\frac{1}{2}$ cycle, $d/\lambda=\frac{1}{2}$, and $\alpha=180°$ (or $0°$) for the forward direction; then G^2 (180°) has a maximum value of $m^2=4$.

The unwanted minor maximums (minor lobes) revealed by Eq. (6-1) cannot be reduced greatly in relative magnitude by changing the number of elements m. If md/λ is large, the factor $F^2(\alpha)$ for the first minor lobe will have a value of about 0.045. Further reductions can be accomplished by increasing the strength of the radiation from the dipoles near the center of the array. The main lobe is broadened somewhat by this procedure, since most of the power is coming from a smaller array.

Although linear arrays are used for certain purposes in the microwave region, a parabolic reflector with a dipole or electromagnetic horn at its focus is the usual arrangement. In fact, even some long-wave systems use paraboloids. Truncated paraboloids with approximately rectangular openings and full paraboloids of revolution with circular aperture are most frequently employed in radar. If the dipole or horn illuminates a circular or rectangular reflector uniformly, there is a plane of constant phase across the face of the reflector[10] and the radiation pattern is almost identical with the diffraction pattern of a plane wave passing through an aperture whose periphery is that of the reflector.

For a uniformally illuminated rectangular paraboloid of length a and width b, the power P per unit solid angle varies according to the well-known relationship,

$$P=P'\left[\frac{\sin^2\pi(a/\lambda)\sin\beta}{\pi(a/\lambda)\sin\beta}\right]^2\times\left[\frac{\sin\pi(b/\lambda)\sin\gamma}{\pi(b/\lambda)\sin\gamma}\right]^2, \quad (6\text{-}3)$$

where β and γ are the angles indicated in Fig. 47, and P' is the power per unit solid angle normal to the opening. For a uniformally illuminated circular opening of radius R, the corresponding relationship is

$$P=P'\left[\frac{2J_1(2\pi\beta R/\lambda)}{2\pi\beta R/\lambda}\right]^2, \quad (6\text{-}4)$$

where β is the angle between the normal to the opening and the radius vector to a point in the radiation field, and $J_1(2\pi\beta R/\lambda)$ is the Bessel function of the first order.

The magnitude of the minor lobe relative to the main lobe from a paraboloid depends on its shape and on the distribution of illumination over its surface, but not on its over-all size. As with arrays, increasing the illumination near the middle of the paraboloid reduces the minor lobes considerably at the expense of a slight broadening of the main beam.

In order that the outward radiation from a dipole antenna at the focus of a paraboloid may be turned toward the parabolic reflecting surface, a small plane reflector such as that in Fig. 39 is frequently used. If the plane reflector is large compared to the dipole, the pattern at the parabola is approximately that given by Eq. (6-2);

[9] See *Principles of radar* (Technology Press), chap. 9.

[10] This assumes that the dipole or horn acts as a point source, a condition that can be fulfilled approximately but not completely.

in other words, the effect is that of an image dipole behind the reflector. In certain cases an actual dipole (parasitic-antenna) may be placed in front of the driven dipole to direct the radiation toward the parabola. Actually the use of parasitic antennas to form a directive radiation pattern without the use of a parabolic reflector was one of the earliest methods used in airborne radar. One parasitic antenna behind the driven dipole and several in front in the direction of propagation at proper spacing form a Yagi antenna. The beam width of a Yagi radar antenna is about 40°.

Gain and beam width.—The concepts of beam width and antenna gain are important in radar. Beam width in a plane is usually taken as the angle between the two directions in that plane for which the power is one-half the maximum power. For a given type of antenna, the beam width in a given plane is approximately proportional to the ratio of the wavelength to the linear dimension of the antenna in that plane, provided the linear dimensions are much greater than the wavelength.

The beam width B (radians) from a linear array of length md is, from Eq. (6-1), approximately

$$B = 0.88\lambda/md. \qquad (6\text{-}5)$$

For a uniformly illuminated rectangular opening of length a, Eq. (6-3) gives

$$B = 0.88\lambda/a. \qquad (6\text{-}6)$$

For a uniformly illuminated circular opening of diameter D, Eq. (6-4) gives

$$B = 1.028\lambda/D. \qquad (6\text{-}7)$$

Thus a circular paraboloid 2 m in diameter will give a beam width of about 3° for a wavelength of 10 cm. An array of eight dipoles spaced one-half wavelength apart gives a beam width of approximately $12\frac{1}{2}$°.

Both angular accuracy and angular resolving power of a radar system depend on the beam width. The former can be increased by such methods as lobe-switching, where the direction of the main lobe is alternately shifted through a small angle from one side of the target to the other, and the antenna is directed so that finally the echoes from the two sides are of equal magnitude. If the switching angle is chosen so that the part of the radiation pattern where the echo amplitude varies very rapidly with angle is used, the angular accuracy can be made a few percent of the beam width. The resolving power is approximately one-half the beam width.

Antenna gain, as defined in Sec. 1, is the ratio of the power per unit solid angle in the center of the main beam to that from a nondirectional antenna radiating the same total power. A single dipole or an open wave guide has a gain of about 1.5. The largest contribution to gain comes from the directive mechanism. For either an array or a uniformly illuminated paraboloid of area A, the gain can be shown to be approximately[11]

$$G = 4\pi A/\lambda^2, \qquad (6\text{-}8)$$

provided $\lambda^2 \ll A$. Thus a circular paraboloid 2 m in diameter would have a gain of about 4000 for a wavelength of 10 cm, whereas an array eight dipoles square would have a gain of about 200.

Antennas as circuit elements.[12]—A single dipole antenna in general will terminate an r-f transmission system in an impedance that has both resistive and reactive components. The resistive component is not associated primarily with the actual resistance of the material of the antenna but with the power radiated by the antenna. The radiation resistance of an antenna is defined as

$$R_r = P/I^2 = V^2/P, \qquad (6\text{-}9)$$

where P is the average power radiated by the antenna, and I and V are the effective values of current and voltage at the input to the antenna.

A thin center-fed antenna less than one-half wavelength long has a capacitive reactance; if greater than one-half wavelength long, it has an inductive reactance. For maximum power transfer from the line through the antenna to free space, the radiation resistance should equal the characteristic resistance of the transmission line and the reactive impedance should be zero. For a thin half-wave antenna the reactance is much more sensitive to change in frequency than for a thicker antenna. Therefore, in practice thick antennas slightly shorter than a half wavelength are used.

When half-wave dipoles are fed by wave guides they may be lined up with the electric field as in Fig. 39. When they are fed by coaxial (or open-wire) transmission lines they may be divided at

[11] Reference 5, pp. 258–263. The gain is, of course, different if the paraboloid is not uniformly illuminated.

[12] A more complete discussion of the impedance of antennas is given in various publications, for example, Hund, *Phenomena in high-frequency systems* (McGraw-Hill), chap. 11.

the center into two quarter-wave sections each of which is fed by one side of the line. Such an antenna is called *center-fed*. The radiation resistance which it presents to the line is from 50 to 75 ohms, which is of the same order of magnitude as the characteristic resistance of transmission lines. The two terminals of the transmission line are sometimes attached part way between the center and ends of a half-wave antenna, giving a higher impedance since the effective current in Eq. (6-9) will then be less near the end of a dipole for the same radiated power. When many dipoles are used, end feeding is common. The two ends of the feed line are connected to adjacent ends of two half-wave antennas. This arrangement presents a very high resistance to the transmission system so that many dipoles can be fed in parallel without terminating the transmission line in too low a resistance.

As a circuit element, a receiving antenna behaves essentially as does a radiating antenna. It acts as a source of r-f energy for the transmission line. For maximum power transfer its radiation resistance must equal the input resistance to the r-f transmission system. As an absorber of radiation a receiving antenna has the same directive pattern that it has as a transmitting antenna. Let us denote by A the effective cross-sectional area, for absorption of radiation, in the direction of maximum directivity. For a single resonant dipole[13] properly matched to its transmission line and parallel to the electric field of a plane wave incident on it,

$$A = 3\lambda^2/8\pi. \tag{6-10}$$

For directive devices of linear dimensions large compared to a wavelength, the effective cross-sectional area A is simply the physical area of the directive device.

Scanning.—In most radars, to search for a target, or to follow it after its location, there must be a way to change the direction of the main beam of the antenna in azimuth and sometimes in elevation. In some radars this is accomplished by turning the whole antenna structure, and in others by variations within the antenna. In the case of an array such as that shown in Fig. 46, the beam can be directed obliquely by introducing a constant phase difference between adjacent radiators. The beam will incline toward the lagging end of the linear array. If this phase difference is continuously changed, the beam will rotate and can be shifted from side to side through a considerable angle. The direction of the beam from a parabolic reflector can be made to deviate from the axis of the paraboloid by moving the source off the axis of the paraboloid while keeping it in the focal plane. In some cases the dipole or horn is mounted off the axis and rotated about it. This arrangement results in conical scanning, which is very useful when the antenna is to follow the target automatically as in fire control. If the target is slightly off the axis of the parabola, the amplitude of the received signal is modulated with the frequency of rotation of the dipole or horn. This modulation is detected by circuits in the radar that rotate the antenna in such a direction as to reduce the modulation to zero.

Propagation.—The radiation patterns of radar antennas have been discussed so far from the point of view of propagation unaffected by reflections from the surface of the earth. This point of view is essentially correct for propagation over land because, for most radar frequencies, the ground waves are rapidly absorbed. Moreover, reflections from the earth are too irregular to produce regular interference effects. Over the sea, reflection is important, unless the elevation angle of the target is large relative to the vertical beam width. When the elevation angle is relatively small, both reflection from the sea and refraction by the atmosphere are usually important.

In the latter case the pattern can be considered as resulting from the original antenna above the surface and an image antenna equally far below the surface. If the sea is considered to be perfectly conducting, for horizontal polarization (electric field horizontal) the image antenna is 180° out of phase with the actual antenna, while for vertical polarization the pattern at small glancing angles predicates an image antenna that is in phase with the actual antenna. Since Brewster's angle for microwaves incident on the sea is large (greater than 80°), the situation for vertical polarization is rather complicated except for glancing incidence.

The presence of image antennas multiplies the free-space pattern by a factor which, if the reflection is perfect, can be obtained from Eq. (6-2)

[13] Reference 5, p. 244.

by making $\varphi = 0°$ for vertical polarization and $\varphi = 180°$ for horizontal polarization. However, the following broad features of the interference patterns for a perfectly conducting sea are clear from physical considerations:

(1) For *horizontal polarization* there is a minimum at the surface and there are maximums for angles above the surface for which the path from the image antenna is an odd number of half wavelengths greater than that from the actual antenna.

(2) For *vertical polarization* there are maximums at the surface and for angles above the surface for which the path from the image antenna is an integral number of wavelengths longer than the path from the actual antenna.

Horizontal polarization rather than vertical is generally used because it seems to give greater visibility of surface targets. The complete explanation of this effect is complicated. Probably it occurs because small water waves and tiny objects on the surface clutter the scope less for horizontal than for vertical polarization, which has maximum intensity at the surface. Moreover, the intensity of horizontally polarized waves increases above the surface (where a portion of the target usually is) while the vertically polarized intensity decreases. Thus, for a fairly smooth sea, horizontal polarization gives a higher ratio of target echo to sea return than does vertical polarization. As the sea becomes rougher the difference between the two decreases.

The angle of the first maximum above the water for horizontal polarization is roughly proportional to the wavelength of the radiation. Thus microwaves ($\lambda < 20$ cm), which give narrower beams and better resolving power, also are propagated closer to the surface of the sea than are longer waves. The angle of the first maximum also is approximately inversely proportional to the height of the antenna above the sea. For this reason, and because the geometric horizon is more distant, search radar antennas are located on the highest part of the ship, while on the seacoast the highest point near the coast is chosen. Before microwaves were extensively used, many successful attacks were accomplished by aircraft approaching the radar at low altitudes, under the first maximum.

If the antenna is a sufficient number (about 100) of wavelengths above the sea, the usual horizon of the radar is roughly the optical horizon. Normally, radar waves are not appreciably reflected by the ionosphere. They are, however, refracted by the atmosphere because of the decrease of density with altitude in a manner similar to the refraction of optical waves, for which the average radius of curvature of their path is four times the radius of the earth. The chief difference here between the behavior of radar waves and optical waves is caused by the dielectric constant of water, which is much larger for radar than for optical wavelengths. Water has a permanent electric moment. For frequencies lower than the infra-red, the water molecules follow the electric field oscillations, whereas for frequencies in the visible region they cannot follow the oscillations very well. The refractive index n of wet air, at radar frequencies, is given by

$$n-1=79\times 10^{-6}\left(\frac{p}{T}+\frac{4800e}{T^2}\right), \qquad (6\text{-}11)$$

where p is the air pressure in millibars, e is the water vapor pressure in millibars, and T is the absolute temperature. The second-order temperature factor results from the increased difficulty of alining the water molecules with the electric field as the thermal velocities increase. This term is, of course, much smaller for visible light.

Although extraordinary refraction occurs when the first term of Eq. (6-11) is large, it does not often produce spectacular effects. On the other hand, the refraction due to water vapor pressure may become very large indeed in such cases as that of a dry wind blowing from land over water. The bottom layers of the air become saturated, whereas higher up the air remains dry. Thus, the refractive index decreases very rapidly with altitude. The radar waves may in such cases have radii of curvature less than the earth's radius. Then they curve down, strike the sea, are reflected and, after starting upward, are again curved down to the sea. The waves may travel up to 100 mi or more over the optical horizon in this fashion, and echoes are obtained at extraordinary ranges. This phenomenon is known as *super-refraction.*

Targets.—The consideration of the cross section of targets is somewhat complicated. Except for resonance effects such as for targets one-half wavelength long, targets intercept the radiation that strikes their surface (σ of SEC. 2). For targets of dimensions small compared to a wavelength the directive factor η (SEC. 2) is very nearly unity. Targets that are large compared to a wavelength can be analyzed into plane and curved

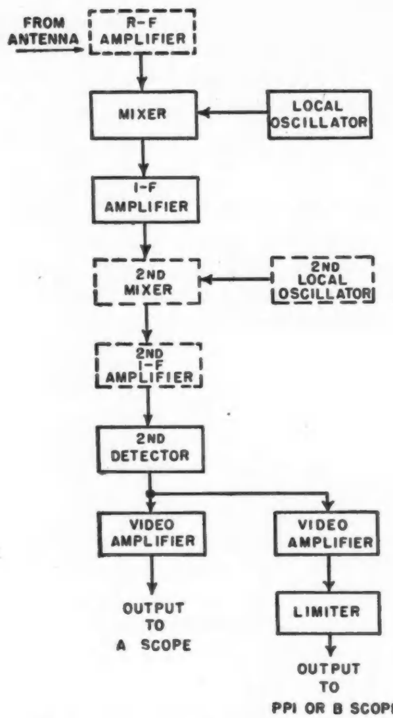


FIG. 48. Block diagram of components of a radar receiver.

surfaces. The factor η is very large for plane surfaces when the radiation strikes them normally; otherwise it is very small. Since most targets move, a plane surface gives fluctuating echoes, which usually average out to a directivity* somewhat greater than unity. A surface of radius of curvature much less than the radar range tends to diverge the radiation whether it is convex or concave. Thus, it has less directivity than a critically oriented plane surface but much more than an oblique plane surface. The directivity of such a curved surface any point of which the radar beam strikes normally is inversely proportional to the average curvature of the surface at the point.

Plane surfaces intersecting at right angles tend to reflect radiation back to its source. In fact, if there are three mutually perpendicular plane surfaces, practically all the radiation will be reflected back toward its source. Such devices are called *corner reflectors* and form the best possible radar targets. Ships and cities tend to give strong echoes because they contain many planes intersecting at right angles.

* In these considerations reflectivity is taken to be 100 percent.

As a result of variations in reflection, refraction and meteorological conditions, no target gives a constant enough echo to be used as a standard for checking the performance of the radar. However, a high Q cavity resonator (echo box) tuned to the transmitter frequency and having an antenna near the radar antenna can be used as a standard target for testing over-all performance of the radar. The energy received by the echo box during transmission is proportional to the energy transmitted. After the transmitted pulse ceases, the amplitude of the resonator oscillations decay exponentially and will meanwhile radiate energy into the radar receiver. If the receiver is tuned to the resonator frequency and hence to the transmitter frequency, the time for the decaying amplitude to fall below the amplitude of the receiver noise (see SEC. 7) is a measure of over-all performance of the radar. "Ringing time" is usually about 30 μsec if the radar is performing satisfactorily.

7. RECEIVERS

As previously indicated, a distant object reflects or scatters the electromagnetic energy falling on it. A very small fraction of this energy, in some instances less than 10^{-18} of the transmitted energy, returns to the antenna and is received by it. The potential differences thus produced in the antenna are of the order of microvolts, while those during transmission are of the order of kilovolts. The received pulses are detected and amplified by a superheterodyne receiver.

Figure 48 is a block diagram of the components of a typical radar receiver. Those components indicated by broken lines on the figure are not used in all radars. For example, r-f amplifiers are not used at wavelengths less than 30 cm. The second mixer, second local oscillator and second i-f amplifier are used only in a few early radar systems, where it is desired to reduce feedback by amplifying at two different frequencies. A qualitative description of a radar receiver has been given in SEC. 1. Therefore, only the more quantitative aspects will be given here.

Receiver Sensitivity

Since the sensitivity of a radar receiver is limited by noise, the nature of this noise is important. Although some noise enters the antenna along with the signal, almost all the noise that appears in the output of a radar receiver originates in the receiver itself, as random fluctuations of current. The important sources of receiver noise are the random voltages due to thermal agitation of electrons in the resistors of the circuits and tube noise, such as random variations in emission and other complicated effects. These important sources of noise have one feature in common:* the equivalent average (rms) noise power P introduced into the input circuit of a receiver by these fluctuations is given by

$$P = K_1 \Delta f, \tag{7-1}$$

where K_1 is a constant depending on the frequency and the type of circuit, and Δf is the receiver band width. The average noise power P limits the smallest received signal that can be distinguished. Experimentally, if f_r is sufficiently large, the ratio of the minimum detectable peak signal power to the average noise power equals

$$P_{r_{\min}}/P = K_2, \tag{7-2}$$

where K_2 is constant and of the order of unity, its exact value depending on the type of scan, speed of sweep, persistence of the screen and ability of the operator. Thus the average noise power P gives the approximate theoretical limit of sensitivity for the receiver.

Receiver sensitivity is stated in terms of a quantity, the noise figure NF, which is the ratio of the equivalent noise power of a receiver to the component of input noise power associated with the antenna resistance. If the receiver is matched to the antenna, the amount of noise power with components in the bandwidth Δf that the antenna radiation resistance supplies to the receiver is $KT\Delta f$, where K is the Boltzmann constant, and T is the temperature in degrees Kelvin. An ideal receiver has a noise figure of unity. Now, since the noise figure is a power ratio, it is generally expressed in decibels as follows:

$$NF = 10 \log_{10}\left[\frac{\text{Average (rms) noise power}}{KT\Delta f}\right] = 10 \log_{10}(K_1/KT). \tag{7-3}$$

Thus the noise figure is independent of bandwidth.

At frequencies below 200 megacyc/sec, noise figures range from 2 to 6 db, while at frequencies of 3000 megacyc/sec and above the values may exceed 15 db. The larger values at higher frequencies are caused by increase of tube noise and decrease of tube input resistance. It can be shown that signals of the order of magnitude of 10^{-16} kw should be detected by a receiver having a noise figure of 10 db.

Pulse Amplification of Rectangular Pulses

Because radar uses pulsed rather than continuous operation, a radar receiver must amplify high frequency pulses of short duration. Therefore, the receiver must respond to a band of frequencies centered on the carrier frequency.

Since the effects of radio-, intermediate- and video-frequency amplifiers on pulse shape are similar, the problem of amplifying high-frequency pulses can best be understood by considering the video amplification of the envelope of the pulses. The pulses at the input of the video amplifier (Fig. 48) may be regarded as rectangular. Periodic rectangular pulses of duration δ and with a time interval $\tau_r[=1/f_r]$ between pulses are shown in Fig. 49.

It is well known from Fourier analysis that a train of periodic rectangular pulses, symmetrical about $t=0$, where t is the independent variable, can be obtained by adding a constant and a large number of cosinoidal waves of various frequencies and amplitudes, so that

$$V(t) = \tfrac{1}{2}A_0 + A_1 \cos 2\pi f_r t + A_2 \cos 2(2\pi f_r t) + A_3 \cos 3(2\pi f_r t) + \cdots + A_k \cos k(2\pi f_r t) + \cdots. \tag{7-4}$$

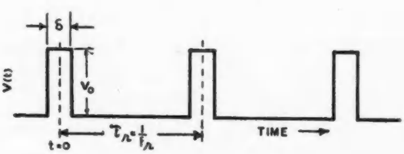


FIG. 49. Periodic rectangular voltage pulses

* Ignition, atmospheric and jamming noise are usually exceptions to the rule. Only the last is ever large enough to be important.

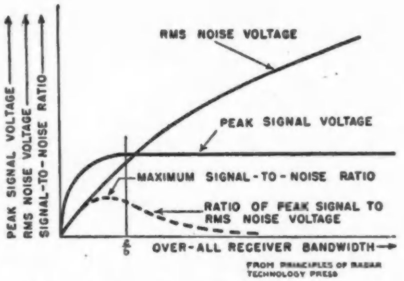


FIG. 50. Receiver band width for maximum signal-to-noise ratio.

Thus the output of the video amplifier may be obtained by adding together the output components of each of the frequency components, since the video amplifier is approximately linear. The video amplifier does not pass a constant voltage and, therefore, the first term in Eq. (7-4) may be neglected.

If some of the Fourier frequencies present in the input pulses are not amplified because of the response of the amplifier, then the pulses will be changed in shape in passing through the amplifier. It can be shown that the frequency response required for a given quality of reproduction is determined only by the pulse duration δ, and, in general, the shorter the pulse duration, the better the high frequency response of the amplifier must be if the quality of reproduction is to be the same. These results described here for video amplification are equally valid for the frequency response of the r-f and i-f amplifiers and for the over-all response of the receiver.

Band Width Requirements

As previously indicated, although the noise figure is independent of the band width [Eq. (7-3)], the average noise voltage in the output of a receiver is proportional to the square root of the receiver band width. But a receiver must pass a large band of frequencies if a reasonably undistorted signal is to be obtained. Thus the band width must be neither so small as to result in loss of signal nor so large as to give excess noise in the output. The use to which the radar is to be put determines very largely the relative importance of the effects of band width on noise and on the shape of the output pulse. For example, in fire control systems the band width is wide in order to insure a steep pulse side for accurate ranging. On the other hand, in search radar the shape of the output pulse is of less importance than the signal-to-noise ratio, for detection of a target at maximum range is the main consideration. If the ratio of peak signal voltage to rms noise voltage in the receiver output is taken as a rough criterion of visibility on an indicator, Fig. 50 indicates that the optimum band width for best detection is given by

$$\Delta f_{\mathrm{opt}} \approx 1.5/\delta. \qquad (7\text{-}5)$$

For a receiver fulfilling the optimum band-width condition, Eq. (7-5), there is a relation between the minimum detectable peak signal power, $Pr_{\min}$, and the pulse duration. Combining Eq. (7-5) with Eqs. (7-1) and (7-2), one obtains

$$Pr_{\min} = K_2P = K_2K_1\Delta f = 1.5K_2K_1/\delta. \qquad (7\text{-}6)$$

Thus the minimum detectable power is inversely proportional to the pulse duration if the band width is properly adjusted. Rewriting Eq. (7-6), one obtains

$$(Pr_{\min})(\delta) = 1.5K_2K_1 \approx 1.5K_1. \qquad (7\text{-}7)$$

The minimum detectable energy is constant since it is approximately $1.5K_1$. But K_1 cannot be reduced indefinitely. Hence, as stated in SEC. 2, it is the energy of a pulse, not its peak power, that determines the detectability of the pulse. Therefore, the expression for $r_{\max}$ [Eq. (2-8)] derived in SEC. 2 here becomes

$$r_{\max} = \frac{1}{2\sqrt{\pi}}\left[\frac{(P_t\delta)(G_tA_r)(\eta\sigma)}{1.5K_1}\right]^{\frac{1}{4}}. \qquad (7\text{-}8)$$

Hence, for the attainment of maximum range the energy of the transmitted pulse must be as large as possible.

Receiver Components

Video amplifiers are resistance-capacitance coupled (Fig. 48) and have a broad band width. In radar the lower half-power frequency is of the same order of magnitude as the pulse repetition frequency f_r, whereas the upper half-power frequency is at least equal to the reciprocal of the pulse duration. Thus the video amplifier usually has a broad enough band width so that it does not appreciably affect the over-all receiver band width. One or two stages of video amplifiers are ordinarily used with a voltage gain of from 20 to 50.

I-f amplifiers used in radars are tuned amplifiers operating at frequencies from 15 to 60 megacyc/sec. They are usually capacitance-coupled and single-tuned, that is, there is only one *RLC* circuit per stage tuned to the frequency to be amplified. Double-tuned amplifiers, in which both the grid and plate circuits of each stage are tuned and in which successive stages are inductively coupled, are occasionally used. The theory of single-tuned i-f amplifiers shows that the magnitude of the maximum gain $|A_n|_m$ of an n-stage single-tuned i-f amplifier is

$$|A_n|_m = [g_m\sqrt{(2^{1/n}-1)}/2\pi C_t(\Delta f)_n]^n, \quad (7\text{-}9)$$

where g_m is the transconductance of the tube used, C_t is the total shunt capacitance of the tuned circuit of a single stage, and $(\Delta f)_n$ is the band width between half-power points on each side of the intermediate frequency. In order to pass a given pulse frequency band, $(\Delta f)_n$ must be approximately twice the Δf discussed in connection with the Fourier analysis (SEC. 7) of a pulse. It is seen from Eq. (7-9) that to obtain a given amplification, more stages must be used if the band width is large than if it is small. Four stages is the minimum number, whereas ten or more stages may be used when the band width is broad (5 megacyc/sec). The i-f amplifier usually limits the over-all receiver band width. The i-f voltage gain ranges from 10^4 to 10^6.

R-f amplifiers use circuits similar to those used in i-f amplifiers except that they are, of course, tuned to give maximum gain at the radiofrequency rather than at the intermediate frequency. Short-circuited transmission lines usually are used instead of tuning coils and, to prevent radiation, are generally of the coaxial type. The stages are usually double tuned, having one tuned transmission line between grid and cathode and another between grid and plate. Special tubes with small electrode spacings must be used to reduce electron transit time. Such tubes are the acorn tubes (types 954 and 956) and the lighthouse tube (type *GL*446). The voltage gain of radar r-f amplifiers ranges from 1 to 20. R-f amplifiers are not used at frequencies much over 1000 megacyc/sec, because their output signal-to-noise ratio is lower than that of a mixer tube or crystal mixer.

A *mixer* is the circuit element that converts the r-f received signal into an i-f signal. Its chief characteristics follow: the output voltage must be a nonlinear function of the input; the input circuit must be tuned to the radiofrequency; the output circuit must be tuned to the intermediate frequency; means must be provided for applying both the r-f signal and the local oscillator output to the input of the mixer without appreciable loss of signal strength. All radar mixers are "single input" mixers, that is, the local oscillator and r-f signals are applied between the same pair of electrodes. For frequencies below about 1000 megacyc/sec an amplifier tube is used as a mixer. Nonlinearity is achieved by biasing the tube at cut-off and adjusting the local oscillator output so that the grid-cathode potential difference is zero at each positive maximum of the sinoidal oscillator voltage. Above 1000 megacyc/sec, silicon or sometimes germanium crystals with a thin tungsten whisker are used. Silicon is a semiconductor, and electrons pass more readily from the whisker to the silicon than in the opposite direction, thereby providing the necessary nonlinearity. Although the i-f output of a crystal per unit r-f input is smaller than for tubes even at 3000 megacyc/sec, the crystal noise is enough smaller to give the crystal the advantage in signal-to-noise ratio. Because of the smallness of the tungsten silicon contact, the crystal is very susceptible to being burned out and to loss of sensitivity owing to damage caused by leakage of the transmitted pulse through the *T-R* box. The latter must therefore provide better protection for crystal mixers than for tube mixers.

Local oscillators may be either of the triode type or of the velocity modulated, or klystron, type. The triodes are used at frequencies less than about 3000 megacyc/sec, whereas velocity modulated tubes are used above 3000 megacyc/sec.

The power output of a local oscillator is small, 100 mw or less, of which only a small part reaches the mixer. Extreme frequency stability is necessary because the receiver band width is almost always less than 1 percent of the radiofrequency and may be 0.1 percent or less. Frequently voltage regulated power supplies and sometimes temperature control are necessary to obtain such stability. A local oscillator must be tunable over a range of from 10 to 40 megacyc/sec in order to compensate for drift in the frequencies of the

transmitter and the local oscillator. Remote tuning is frequently desirable and, in such cases, electric tuning is used if possible.

Triode local oscillator circuits are similar to triode transmitter circuits in being reducible on paper to Colpitts-type oscillators with feedback through the interelectrode capacitances. For frequencies of a few hundred megacycles per second, conventional tuned circuits are used, whereas, for higher frequencies, sections of coaxial transmission lines are used. The tubes in most frequent use are the type 955 acorn, types 316A and 368A doorknob, and type GL446 lighthouse.

Radar velocity-modulated local oscillators operate on the principle of the two-cavity klystron,[14,15] but differ from it in that: (i) low Q* cavities are used, since large power output is no object and breadth of tuning with nearly constant power output is important; (ii) only one cavity and one pair of grids are used instead of two. When the electron beam from the cathode passes the pair of grids, the cavity field between the grids acts as a "buncher"; that is, the electrons that arrive at one moment are changed in velocity a different amount than those that come later. In place of the second cavity and grids there is a repeller electrode which repels the electrons back toward the original cavity and grids. Those electrons with the higher velocities take longer to return and hence can be made to arrive back simultaneously (bunched) with the later, slower electrons. If a bunch passes back between the grids at a time equal to $(n+\frac{3}{4})$ periods later, where n is an integer, then the cavity electric field is a maximum in the direction of motion of the electrons, some of the electron energy will be given to the oscillating field of the cavity and oscillations will be sustained. Such a tube is known as a reflex velocity-modulated tube.

The frequency of a reflex tube is roughly set by mechanically tuning the cavity. Varying the repeller-cathode potential difference gives a fine frequency control which can be used for remote tuning. In some circuits the local oscillator is automatically tuned to the transmitter frequency by using the output of a discriminator circuit to control the repeller voltage. In this way, compensation is obtained for any change in transmitter frequency.

The detector in a radar rectifies the output of the last i-f stage and by-passes the intermediate frequency component to ground. Its output is therefore a rectified envelope of the input i-f pulse. The most frequently used detector is a diode. The input is usually connected to the cathode so that the output pulses from the plate are negative. Thus, if only one stage of video amplification is used there is no possibility of a strong signal driving its grid potential above that of the cathode and developing a large "signal" bias which would block the video amplifier for a considerable time after the signal has passed.

8. CATHODE-RAY TUBE INDICATORS

The position of a target is described completely if its range, azimuth and elevation relative to the radar are known. In search radar it is desired to present simultaneously the coordinates of a number of targets. In general, only two coordinates—which two depending on the particular application of the radar—are displayed on the cathode-ray screen.* However, a rough indication of a third coordinate is obtained in some types of presentation. Two of them are always presented with good precision.

Two methods of applying the receiver output are employed in cathode-ray tubes used in radar:

(i) Deflection, or position, modulation, in which the sweep voltages are applied to one pair of deflection plates and the receiver output is applied to the other pair.

(ii) Grid, or intensity, modulation, in which the receiver output is applied to the control grid of the cathode-ray tube, while sweep voltages are applied to both sets of deflecting devices. Thus the intensity of the electron beam varies with the strength of the receiver output.

Cathode-ray tubes are classified in accordance with the method used to focus and deflect the electron beam. In radar two types of cathode-ray tubes are used:

(i) *Magnetic tubes.* In these tubes magnetic fields are used both to focus and deflect the electron beam. The tubes are suited to those applications in which a well-focused beam of high current density is desired so as to give an

[14] Reference 2, pp. 329–340.
[15] Ginzton and Harrison, *Proc. I.R.E.* **34**, 97*P* (1946).
* Q is defined as the ratio of the energy stored in a system to that lost per cycle.

* For a description of cathode-ray tubes see Millman and Seeley, *Electronics* (McGraw-Hill), p. 63 ff; Maloff and Epstein, *Electron optics in television* (McGraw-Hill), p. 163 ff.

extremely high screen illumination. Such tubes are therefore suitable for intensity modulation.

(ii) *Electrostatic tubes.* Here electric fields are used to focus and deflect the electron beam. Because they are lighter than magnetic tubes, they are used for intensity modulation wherever weight is an important factor. They are generally used in all applications except where high screen illumination is required. The linear sweep circuits (SEC. 9) are simpler than for magnetic tubes, especially for fast sweep traces that are linear with time.

Electrostatic Cathode-Ray Tube

A schematic diagram of an electrostatic cathode-ray tube is shown in Fig. 51. The *electron gun* consists of: the source of electrons, which is an indirectly heated cathode; a control grid; a first, or focusing, anode; and a second, or accelerating, anode.

The voltage applied to the control grid limits the number of electrons reaching the screen, and the voltages applied to the first and second anodes accelerate the electrons. Convergence of the beam of electrons depends upon the ratio of the voltages applied to the first and second anodes. When the electron paths meet at a point on the screen, the beam is said to be in focus. In practice, focus is usually obtained by adjusting the first anode potential. The cathode, control grid and the first and second anodes are cylinders. The axis of the cylinders are coincident with the axis of the tube. The cathode is oxide-coated and heated by means of a non-inductively wound filament. The baffle plates shown in the first and second anode cylinders in Fig. 51 intercept electrons that follow paths too divergent for good focusing.

As shown in Fig. 51, just beyond the second anode two sets of plates are mounted at right angles to each other, one set being in the vertical plane and the other in the horizontal. The electron beam is deflected by the electric fields due to potential differences between the vertically and between the horizontally deflecting plates. To prevent the beam from striking the plates, when large potential differences are applied, the plates are made parallel for a part of their length and divergent for the remainder, as shown in the diagram. Since the intensity of the spot depends on the velocity of the electrons at the screen and the number per unit time incident on unit area, the spot intensity may be increased by using a higher second-anode potential, but this results in decreased deflection sensitivity. The high voltage auxiliary anode shown in Fig. 51 is used in some tubes to permit higher accelerating voltages where greater screen illumination is required without loss in deflection sensitivity. The auxiliary high-voltage anode may be operated at a potential 30 to 100 percent higher than the second-anode potential.

Most electrostatic tubes used in radar employ fluorescent screens of medium persistence emitting green or white light, depending on the type of phosphor used. When long persistence is required, a screen known as a *cascade* is used. It is composed of two actual layers: a blue (flash) layer, which is the one nearer to the electron gun; and an amber or yellow layer. The latter layer is made of zinc sulphide or cadmium sulphide, with copper as an impurity, which acts as an activator. The blue layer is zinc sulphide with silver as an activator. The excitation of the amber layer by the light from the flash layer results in long persistence. The light output per unit electron current varies with the square of the phosphor potential.[16] When long persistence is required, an electrostatic tube with an auxiliary high voltage anode is used since the incident electrons must have an energy of about 4000 ev to produce satisfactory results.

The electrons reaching the screen and giving up their energy in producing luminescence must be removed. The incident electrons tend to produce secondary emission, and some of these secondary electrons have sufficient velocity to reach the second anode or the Aquadag coating and thus return to the power supply. The remaining secondary electrons fall back onto the

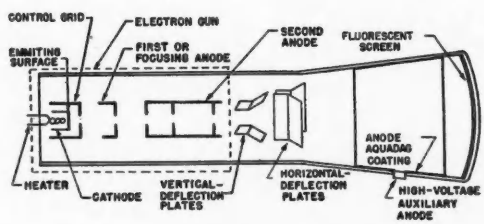


FIG. 51. Electrostatic cathode-ray tube.

[16] W. B. Nottingham, *J. App. Physics* 10, 73 (1939).

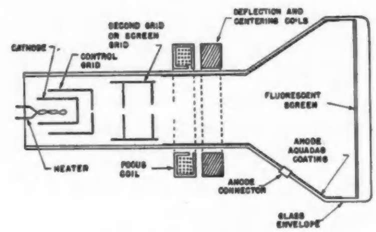


FIG. 52. Magnetic cathode-ray tube.

screen. When the number of secondaries returning to the power supply equals the number of primaries incident on the screen, an equilibrium condition is set up and the screen is maintained at constant potential. In radar indicators, equilibrium is reached when the screen is at a potential of about 100 v below that of the collector anode.

Magnetic Cathode-Ray Tube

Figure 52 is a diagram of a magnetic cathode-ray tube. The section of the tube that includes the cathode heater and control grid operates in the same manner as the corresponding parts of an electrostatic tube. The second, or screen, grid shown in Fig. 52 does not focus the beam but reduces the field, in the region between the cathode and control-grid, owing to the high anode-cathode potential difference, thus reducing the potential difference from cathode to control grid needed to cut off the beam current. The beam is

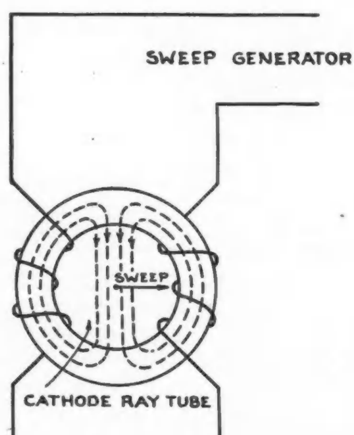


FIG. 53. Single coil deflection yoke.

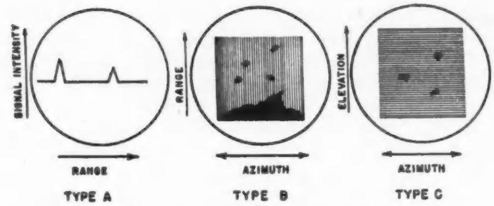


FIG. 54. Types A, B and C presentations.

focused by means of a coil of many turns wound on a soft-iron ring with an annular air gap, as shown in Fig. 52. An adjustable direct current in the coils sets up a strong nonuniform magnetic field at the center of the coil. The focusing coil is usually mounted so that it may be tilted to center the beam on the screen. In some cases a permanent magnet may be used instead of the focusing coil.

The electron beam is deflected by the magnetic field produced by a current in deflection coils (Fig. 52). The coil assembly is called the deflection yoke. The number of coils and their arrangement depend on the kind of deflection and type of presentation desired. Figure 53 shows a deflection yoke with a single coil, designed for use with *PPI* presentation.

A screen of long persistence is usually used in magnetic cathode-ray tubes of radar indicators. Magnetic tubes are well suited to Type *B* and *PPI* presentations, where a bright well-focused spot is desired. With these types the antenna moves rapidly and radiation is incident on the target during only a short part of the scanning cycle. Thus the average intensity would be low, unless it is high while the echo is being received.

Types of Presentation

Because the information provided by the radar echo must be interpreted as rapidly and accurately as possible, many special types of data presentation have been developed to meet the wide variety of tactical and operational situations in which radar is used. However, only those types commonly used, such as the *A*, *R*, *B*, *C*, *P* (*PPI*) and *J* types of presentation, will discussed here.

In type *A* presentation, deflection modulation is used. A horizontal trace, or time-base, as shown in Fig. 54, is produced by a sweep circuit

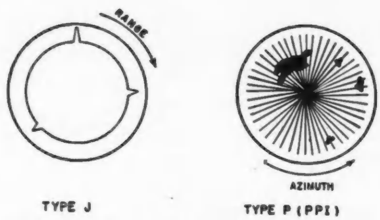


FIG. 55. Types J and $P(PPI)$ presentations.

(SEC. 9). The signal appears on the screen of the cathode-ray as a vertical deflection from the time-base, its abscissa being proportional to the range. By changing the sweep speed, this type of presentation may be modified so as to magnify any small portion of the total range. This is known as R-scan.

In the B-type of presentation the signal intensity-modulates the electron beam, and targets appear as bright spots on the screen. This scan displays, in rectangular coordinates, range *versus* azimuth angle (Fig. 54). In C-type presentation, elevation *versus* azimuth angle is displayed.

In J-type presentation a circular sweep is used. Targets are indicated by radial deflections along the circular trace, as shown in Fig. 55. A deflection caused by the transmitted pulse appears at the top of the screen, and the distance along the circumference of the trace from this deflection to an echo indication is proportional to the target range.

As stated in SEC. 1, type P gives a maplike presentation. The time base starts from the center of the screen and moves radially outward, so that range is measured radially, and azimuth is measured angularly. In Type P, as with most of the intensity-modulated indicators, magnetic cathode-ray tubes with a screen of long persistence are used.

9. SWEEP AND RANGING CIRCUITS

The radar indicator contains, in addition to the cathode-ray tubes, a circuit which, when triggered by the transmitter, determines the rate and duration of the sweep; a second circuit that allows electrons to reach the screen only during the sweep; and a third circuit that furnishes means for accurately measuring the range.

Two general types of sweep circuit are used—electrostatic and magnetic. Despite wide variations in the sweep circuits of different radars, two general statements can be made. First, the time duration of the sweep is nearly always determined by a single-pulse multivibrator[17] which is triggered by a pulse from the transmitter (Figs. 1

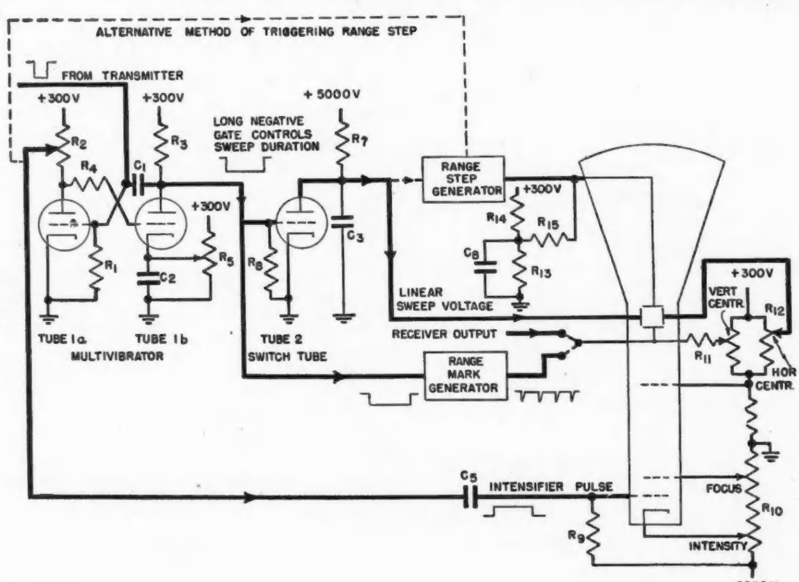


FIG. 56. A typical sweep and ranging circuit for a type A presentation.

[17] Puckle, *Time bases* (Wiley), p. 50.

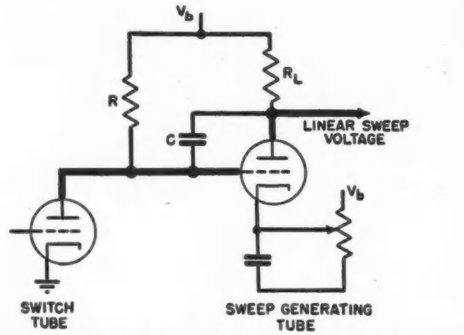


FIG. 57. Linear sweep generating circuit using a low-voltage power supply.

and 56). Second, the output of this multivibrator nearly always cuts off a switch-tube (Fig. 56) which starts the sweep generator. The sweep is generated only while the switch-tube is cut off. Thus the duration of the multivibrator pulse determines the duration of the sweep.

In Fig. 56 tube 1*a* of the multivibrator is normally conducting. Tube 1*b* is normally cut off because R_5 is adjusted so that the cathode potential of the tube is substantially above the grid potential, which is equal to the plate potential of tube 1*a*. The short negative trigger pulse from the transmitter cuts off tube 1*a* and turns on tube 1*b*. In the same manner as described in SEC. **3**, the multivibrator remains in this state for the time required for C_1 to discharge through R_1 and tube 1*b* sufficiently to raise the grid potential of tube 1*a* above cut-off and return the multivibrator back to its original state.

Electrostatic Sweep Circuits

The sweep, which begins when the switch tube is cut off by the multivibrator, must be essentially linear. Before the start of the sweep, tube **2** in Fig. 56 is conducting strongly, and the potential difference across C_3 will be very small if the resistance R_7 is very large. When the switch-tube is cut off, the capacitor C_3 begins to charge exponentially through R_7. If the applied voltage is large (5000 v) compared to the sweep voltage amplitude desired (200 to 300 v), C_3 will charge for only a small fraction of the *RC* time constant and the rise of voltage will be essentially linear.

An alternative method which eliminates the use of high voltages is shown in simplified form in Fig. 57. When the switch-tube is cut off, its plate potential (and hence the grid potential of the sweep generating tube) tends to rise. If the upper side of capacitor *C* were at a fixed potential, the potential of its lower plate would rise exponentially in the same manner as that of the upper plate of C_3 in Fig. 57. However, a rising grid potential on the sweep generating tube (operating class *A*) reduces the plate potential and thus retards the rise of grid potential. The result is an increase in effective time constant for charging *C* and, more important in this case, an increase in linearity, which is the equivalent of having a charging voltage much larger than V_b. Analysis by class *A* equivalent circuits shows that the time constant τ for the charging of *C* is approximately

$$\tau = RC[\mu+1-\mu r_p/(r_p+R_L)] \approx (A+1)RC, \quad (9\text{-}1)$$

where $A=\mu R_L/(r_p+R_L)$ is the gain of the sweep generating tube, and μ and r_p are the amplification factor and the dynamic resistance, respectively, of the sweep generating tube. The equivalent charging voltage is

$$V_b' = V_b(\mu+1)\frac{R_L}{(r_p+R_L)} \approx A\,V_b. \quad (9\text{-}2)$$

In practice, instead of connecting the capacitor *C* directly to the plate of the sweep generating tube a cathode follower or two additional stages of amplification are inserted. The factor *A* of Eqs. (9-1) and (9-2) then must be the over-all gain of the series of tubes.

The approximately linear sweep voltage is connected to one of the horizontally deflecting plates in Fig. 56 while the potential of the other deflecting plate is held constant. The spot produced by the electron beam is thus deflected across the face of the cathode-ray tube at a constant rate. At the same time a rectangular

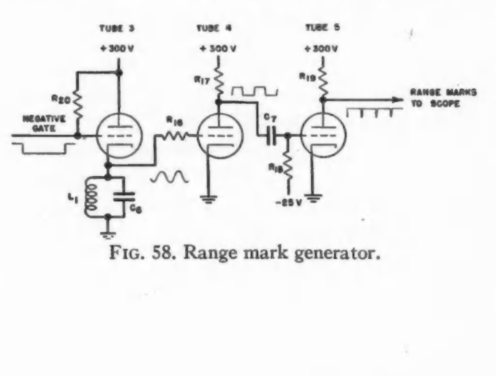


FIG. 58. Range mark generator.

positive pulse from the multivibrator has raised the control grid potential of the cathode-ray tube above cut-off for the duration of the sweep, thus causing the spot to be visible as it travels across the face of the tube. A typical power supply for the tube is shown in Fig. 56, with intensity, focus, and horizontal and vertical centering adjustments. In some radars the sweep voltage is applied to both horizontally deflecting plates through a push-pull amplifier.[18] Use of such an amplifier gives good linearity with a much lower voltage supply for the sweep generator both because of the amplification and because the change in the potential of each deflecting plate is smaller.

Ranging circuits.—Very rough range measurements can be made by ruling a scale on the face of the cathode-ray tube. This method, however, involves errors due to changes of circuit constants with temperature, humidity, tube replacement and so forth. A better method is to put electronic range marks on the screen. The range mark generator (Fig. 56) serves this purpose. A range mark generator consists of an oscillator (usually sinoidal) of accurately predetermined frequency, and suitable shaping circuits to change the sine waves into a train of sharp voltage impulses which are applied to the vertically deflecting plates of the oscilloscope and form uniformly spaced electronic *range marks* on the screen. A very simple circuit for this purpose is shown in Fig. 58. The negative "gate" from the multivibrator cuts off tube **3**. The energy stored in the inductance L_1 starts slightly damped oscillations in L_1 and C_6, with the cathode potential of tube **3** initially going negative. Tube **4** is an overdriven amplifier which changes the sine wave to a square wave (SEC. **3**). C_7 and R_{18} form a peaking circuit (SEC. **3**) that produces positive and negative peaks at the grid of tube **5**. Since this tube is biased below cut-off, only the positive peaks are amplified, and they appear as sharp negative pulses at the plate of tube **5**. When applied to the lower vertically deflecting plate of the oscilloscope, they produce uniformly spaced upward "spikes" (range marks) on the type-*A* presentation. If both echoes and range marks are showing on the oscilloscope simultaneously, ranges can be estimated by interpolation. In some radars quick starting oscillators, which are sometimes crystal controlled, are used instead of the so-called "ringing circuit" of Fig. 58.

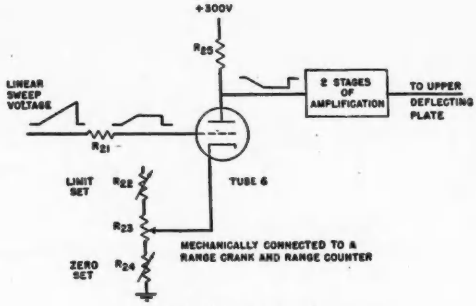


FIG. 59. Range step circuit.

Greater accuracy is possible if a range step generator, shown in a block in Fig. 56, is used. One type of range step circuit is shown in Fig. 59. The cathode potential of tube **6** is positive and adjustable. As the sweep voltage rises, the grid potential of tube **6** reaches cut-off at a sweep voltage (sweep time) determined by the cathode potential. As the grid potential rises above cut-off, the plate potential drops sharply. This drop is steepened by two stages of amplification and is applied to the upper vertically deflecting plate of the oscilloscope. The result is a sharp vertical drop, or *step*, in the time base (Fig. 1 and SEC. **1**). The position of this step can be adjusted by means of R_{23} in Fig. 59. To obtain the range of a target the step is set on some particular part of the echo by means of a crank connected to R_{23}, which is a potentiometer specially constructed to give linearity in the region of calibration. Turning the crank automatically turns a veeder counter which indicates the range in yards or miles. The range counter is made to read correctly against certain specified range marks by adjusting R_{22} and R_{24}.

The range step circuit shown in Fig. 59 is only one of several types that are used. Some of the other circuits are triggered directly by the multivibrator of Fig. 56 and generate a linearly variable delay independently of the sweep voltages. In other radars the step is generated from sine waves of an oscillator that is either synchronous with or started by the trigger pulse of

[18] Reference 17, chap. 8.

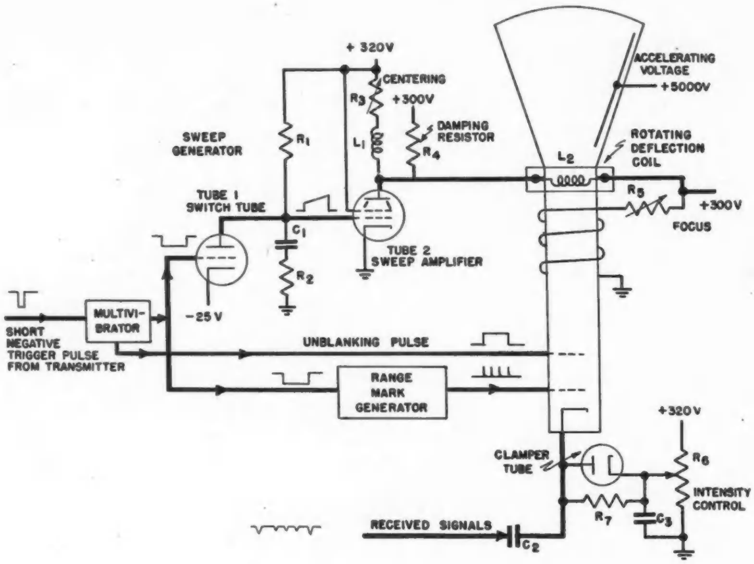


FIG. 60. A simple magnetic sweep and ranging circuit for P type presentation.

the transmitter. In such radars the step is moved by shifting the phase of the sine waves through an accurately known variable angle.

Magnetic Sweep Circuits

A magnetic sweep circuit is shown in Fig. 60. As in the circuit of Fig. 56, a negative trigger pulse from the transmitter operates a single-pulse multivibrator that produces a negative gate for the switch tube of the sweep generator and for the range mark generator and also a positive "unblanking" pulse for the cathode-ray tube.

As stated in SEC. **8**, the electron beam in a magnetic tube is deflected by a transverse magnetic field. For the time base to be linear, the magnetic induction must change linearly with time, so that if the permeability of the magnetic circuit is independent of the magnetic induction, there must be a linearly varying current in the deflecting coil (Fig. 60) during the sweep. This current is controlled by tube **2**, which is a beam-power tube. Figure 61 shows the equivalent circuit of this tube, the deflecting coil and the damping resistor.

Since the current in the coil must be

$$i = kt, \tag{9-3}$$

where k is a constant depending on the rate of sweep desired, it follows from the equivalent circuit that

$$\mu e_g = 0$$

for t less than zero, and

$$\mu e_g = k\left[\frac{L_3(r_p+R_4)}{R_4} + \frac{r_p+R_3(r_p+R_4)}{R_4}t\right] \tag{9-4}$$

for t greater than zero. Thus the sweep signal voltage applied to the grid of the sweep amplifier must be trapezoidal in form, consisting of a square wave of amplitude

$$A = kL_3(r_p+R_4)/\mu R_4 \tag{9-5}$$

and a sawtooth of slope

$$S = \frac{k}{\mu}\left[\frac{r_p+R_3(r_p+R_4)}{R_4}\right]. \tag{9-6}$$

Of the several methods used to generate such a trapezoidal wave, the one shown in Fig. 60 is the most common. It is similar to the sweep generator of Fig. 56 except for the addition of resistor R_2. As before, the time constant $C_1(R_1+R_2)$ must be long compared to the sweep duration, and the voltage of the source must be large compared to the output trapezoid if the square wave

is to be of constant amplitude and the sawtooth wave is to be linear. Under these conditions the square wave has an amplitude

$$A=(320V)R_2/(R_1+R_2), \qquad (9\text{-}7)$$

while the sawtooth has a slope

$$S=(320V)R_1/(R_1+R_2)^2C. \qquad (9\text{-}8)$$

By substituting Eq. (9-8) in (9-6) and Eq. (9-7) in (9-5) and specifying the fraction of a time constant to be used, the approximate values of R_1, R_2 and C can be found.

The grid of the sweep amplifier is just above cut-off between sweeps. If the spot is to be at the center of the screen at the start of a sweep, there must be no current in the coil L_2 (magnetic induction zero). The quiescent current of the sweep amplifier must therefore reach the tube by another path so designed that there is no potential difference across the coil L_2 (Fig. 60); R_3 and L_1 in Fig. 60 comprise such a path. The resistance of R_3 is fairly low and can be adjusted to make the potential difference across it, due to the quiescent current, just 20 v. Inductor L_1 is very large and serves to keep this current constant during the sweep so that all variational currents flow through either R_4 or L_2. The damping resistor R_4 damps out oscillations between sweeps in the resonant circuit formed by the coil L_2 and stray capacitance.

The type of coil shown in Fig. 60 is called a "rotating coil" because it is mounted on ball bearings and is physically rotated about the neck of the cathode-ray tube with the same angular velocity as the antenna. Thus the radial time base rotates with the antenna and gives a maplike presentation on the screen, the center of the map representing the location of the radar.

On nearly all magnetic presentations the received echo appears as an intensified spot on a dark screen. Hence the spot must be cut off completely between sweeps and must be "unblanked" during the sweep just enough so that in the absence of signals a few faint spots of noise appear. The positive unblanking pulse from the multivibrator and the intensity control enable these conditions to be fulfilled.

The amplified received signals (which are shown negative in Fig. 60) are impressed on the cathode of the cathode-ray tube, and the spot is thus intensified during the transmitted pulse (part of which enters the receiver) and during each received echo pulse. The intensity control is connected to the cathode through a "clamping diode." Without this diode, large negative signals might charge C_2 through R_7 sufficiently to bias the cathode positively and hence make it impossible for succeeding signals to intensify the spot enough to be visible; C_2 is rapidly discharged through the diode after such a signal.

Range marks formed in a manner similar to those used with electrostatic tubes may be applied to the control grid as sharp positive voltage pulses. As explained in SEC. 1, these range marks appear as intensified circles. For accurate ranging a movable circle may be provided in a manner similar to that used in providing a movable step on the type-A presentation.

The most common variations from the sweep circuit of Fig. 60 are of the following types. A fixed magnetic yoke with two perpendicular pairs of coils wound on it can be made to give a slowly rotating radial time base by varying the amplitude of the sawtooth current in one pair of coils sinoidally and varying that in the other pair cosinoidally. Two pairs of push-pull sweep amplifiers are usually used. The grid-cathode potential differences for these amplifiers are obtained from orthogonally spaced secondaries of a rotary transformer whose single winding primary rotates with the antenna. The trapezoidal sweep voltage is then applied to the primary of the rotary transformer.

Another variation is the addition of two extra stages of sweep voltage amplification between the sweep generator and the beam-power tube. Usually negative feedback proportional to the total beam-power tube current is returned to the input. With this arrangement the beam-power tube is usually cut off between sweeps, so that the centering circuit is unnecessary.

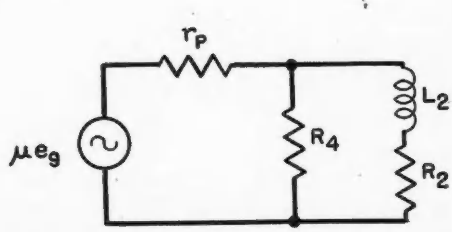


FIG. 61. Equivalent circuit of sweep amplifier or driver tube; μ is the amplification factor of the tube, and e_g is the sweep signal voltage applied to the grid of the tube. Typical values of the circuit constants are $L_2 \approx 60$ mh, $R_2 \approx 50$ to 100 ohms, $R_4 \approx 15{,}000$ ohms. A type 807 tube is frequently used.

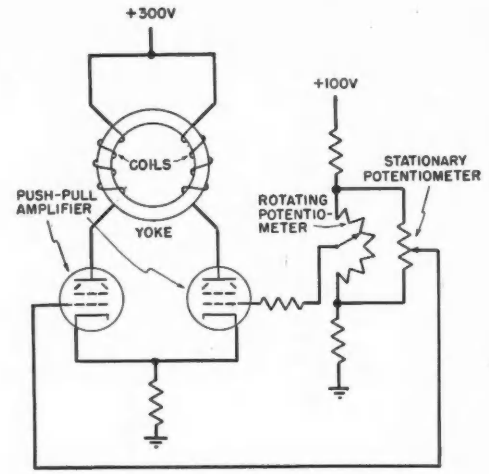


FIG. 62. Magnetic azimuth sweep for type B scan.

In magnetic type-B scans the range sweep generator is similar to that of the P-type but does not rotate. The range time base is always vertical and is moved from left to right as the antenna moves from left to right. The horizontally deflecting coils must carry currents proportional to the angular deviations θ of the antenna from some zero direction. The same equations [Eq. (9-3) through (9-7)] apply as for the time base except that the sweep is so slow (not more than a few sweeps per second), that the effect of the coil inductance is negligible. Thus no square wave is necessary, and potentials

$$v = k_1\theta, \tag{9-9}$$

where k_1 is a constant, must be supplied to the grids of two push-pull beam-power tubes as in Fig. 62. Such potentials can be obtained from a suitably wound potentiometer which has a contact that moves with the antenna.

Electrostatic B scans use a similar potentiometer and push-pull amplifier. In this case, however, the amplifier is a voltage amplifier whose output is connected to the horizontally deflecting plates of the cathode-ray tube.

10. REMOTE CONTROL AND INDICATION

In many radar systems it is necessary to transmit angular data and "orders" over considerable distances. For example, in the search radar of Fig. 1, SEC. 1, both the time base on the *PPI* indicator and the rotation indicator must show accurately the antenna position. Moreover, the radar operator, by means of a hand crank or motor, must be able to transmit to the antenna desired position "orders" which will be accurately followed by the antenna. In fire-control systems under complete radar control, the range and antenna-position data must be transmitted to a computer, and the resulting gun-aiming data must be transmitted from the computer to the gun-aiming mechanism.

Two general classes of remote-control and indication systems are in general use. These are: *synchro-systems*, which transmit angular data without torque amplification and furnish control signals for servo-mechanisms; and *servo-mechanisms*, which, using torque amplification, are able to accurately position bodies that have a large moment of inertia or considerable rotational friction.

Synchro-systems.—Synchro-systems are combinations of two or more of the following five basic types of units: the *synchro-generator*, the *synchro-motor*, the *differential generator*, the *differential motor*, and the *control transformer*. Details of the theory and construction of these units are discussed elsewhere.[19] Only their general characteristics will be given here.

A synchro-generator and a synchro-motor are similar in construction, each having three windings spaced 120° apart on the stator and a single winding on the rotor. The rotor winding of the synchro-generator is connected to a 115-v, 60-c/sec alternating source. The generator acts as a transformer with one primary (rotor) winding and three secondary (stator) windings. The voltage induced in each of the stator windings depends on the angular position of the rotor. The input to the generator is a mechanical position signal applied to its rotor, while the output consists of three electric signals from the three stator windings. The synchro-motor, or follower, on the other hand, receives from a generator three electric signals at its three stator windings. If connections are properly made, these three signals set up an alternating magnetic

[19] L. A. MacCall, *Fundamental theory of servo-mechanisms* (Van Nostrand); also, reference 9.

field whose orientation is the same as that of the generator armature. The rotor of the motor, which is connected to the same 115-v, 60-c/sec alternating source as the rotor winding of the generator, will aline itself with the stator field and thus reproduce the angular position of the rotor of the generator.

The differential generator and the differential motor each have three windings spaced 120° apart on both stator and rotor. A differential generator receives at its three stator (or rotor) windings electric position signals from a synchro-generator. A differential generator also receives a mechanical position signal on the shaft of its rotor and generates in its three rotor (or stator) windings an electric position signal equal to either the sum or the difference of the two input signals, depending on the connections. A differential motor, on the other hand, receives electric position signals at both the rotor and stator windings. The resulting rotor position will be either the sum or difference of the two electric signals, depending on the connections.

A control transformer normally has three stator windings, which receive an electric position signal, and one rotor winding. The amplitude of the alternating voltage generated in the rotor winding varies as the sine of the angle by which the rotor winding deviates from the normal to the resultant alternating field of the stator. A desired angular position is usually connected to the stator windings, whereas the rotor position represents actual angular position. For small angular differences an alternating error signal is generated in the rotor winding that is directly proportional to the difference between stator and rotor positions. This error voltage is amplified and then utilized by servo-mechanisms to turn the control transformer and the load shaft in a direction to reduce the error.

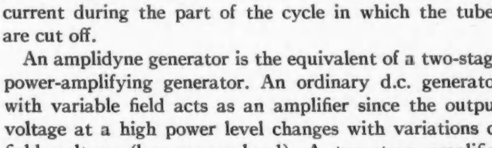
The use of synchro units in a complete system will be discussed in connection with Fig. 64.

Servo-mechanisms.—In radar the servo-mechanism used to drive the load according to the error signal from the control transformer is usually of the amplidyne type, although several other systems are also used. An amplidyne servo-mechanism consists of three basic units, a *control amplifier*, an *amplidyne generator* and a *servo-motor* to drive the load. The control amplifier (electronic) amplifies the error signal of the control transformer. The output of the amplifier supplies the control field of the amplidyne generator. The latter supplies power to a reversible, adjustable-speed servo-motor, causing this motor to reduce the error signal of the control transformer.

The basic circuit of a control amplifier is shown in Fig. 63. The output of the control transformer is an alternating voltage whose phase reverses as the sign of the error in the rotor position changes. The plate supply of the tubes is alternating and from the same power source as the alternating voltage of the control transformer. When no error signal is applied to the grids of the tubes, the currents through the two control fields of the amplidyne are equal and produce no net field since the fields of the two coils oppose each other. However, when an error signal is applied to the grids, it causes one tube to conduct more and the other less during the half-cycle in which the plate potentials of the tubes are positive; the result is a pulsating field in the amplidyne. The direction of the field is dependent on the sign and the magnitude on the amplitude of the error signal. The pulsations can be much reduced by connecting, across the field coils, resistors that carry current during the part of the cycle in which the tubes are cut off.

An amplidyne generator is the equivalent of a two-stage power-amplifying generator. An ordinary d.c. generator with variable field acts as an amplifier since the output voltage at a high power level changes with variations of field voltage (low power level). A two-stage amplifier results from using the output of the first generator to excite the field of a second generator. In an amplidyne both stages are combined in the same armature. The output of the first generator is short circuited. The resulting large armature currents produce a strong magnetic field at right angles to the first set of brushes; this field will generate a still larger voltage between a second set of brushes oriented 90° from the first set. Thus two stages of amplification are achieved on the same armature. Since considerable power and current are to be drawn from the second set of brushes and since this will produce a field through the armature opposing the original exciting field (negative feed-back), a compensating winding is connected in series with the load so that the net armature field of the load current is zero.

The servo-motor used with an amplidyne may be a d.c. shunt motor, a universal shunt motor, or a permanent



FROM PRINCIPLES OF RADAR TECHNOLOGY PRESS

FIG. 63. Basic circuit of amplidyne control amplifier.

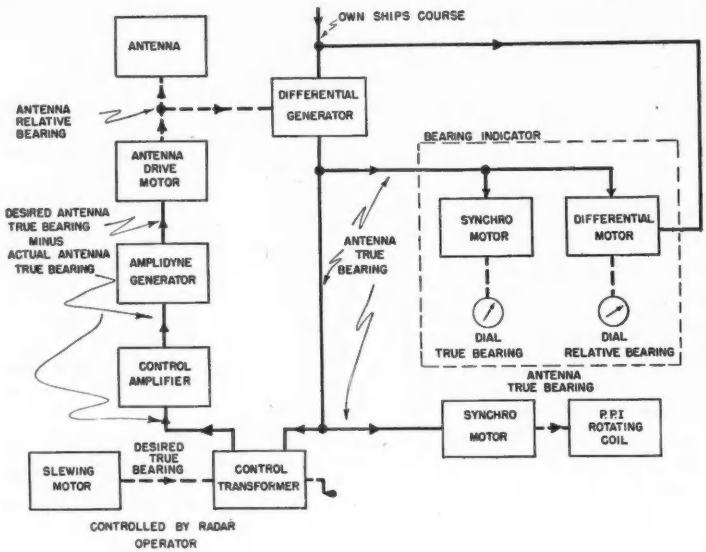


FIG. 64. A possible servo-mechanism for azimuth (bearing) control of a radar antenna on a ship, together with a synchro system for transmission of azimuth data to the bearing indicator and *PPI* oscilloscope of Fig. 1 (SEC. 1). Solid lines indicate paths of electric signals; broken lines, paths of mechanical signals.

magnet d.c. motor. Whatever type is used, the magnitude of the exciting field is constant. The voltage applied to the armature is proportional to the shaft error which the motor must correct and is in the proper direction to reduce the error. In servo-mechanisms employing a.c. amplification, a split-phase, induction type servo-motor is usually used.

The block diagram of a typical complete servo-synchro system of a shipborne search radar is shown in Fig. 64.

The servo-system will be considered first. The rotating shaft of the antenna turns the rotor of a differential generator, and the stator of the generator receives an electric position signal (own ship's course or ship's bearing) from a synchro-generator which is connected to the ship's gyro-compass. The sum of the mechanical signal on the rotor (antenna relative bearing) and the electric signal on the stator (ship's bearing) is the antenna true bearing, which constitutes the electric output signal from the rotor of the differential generator. This antenna true-bearing signal is connected to the three stator windings of a control transformer, the rotor of which can be turned by the radar operator either with the slewing motor or by hand. The rotor position constitutes the desired true bearing of the antenna. The output of the control transformer is the difference between the desired antenna true bearing and the actual true bearing. This error signal is connected to the control amplifier, which in turn excites the field of the amplidyne generator. The latter generates a voltage of the proper magnitude and direction to reduce the error and to make the antenna bearing coincide with the desired bearing.

The *synchro*-system drives the bearing indicator and the *PPI* rotating coil (Fig. 1, SEC. 1 and Fig. 64). The antenna true-bearing signal drives the rotating coil of the *PPI* indicator by means of a synchro-motor. The *PPI* pattern therefore changes only with the position of the ship and not with its course.

The antenna true-bearing signal also drives a true-bearing dial in the bearing indicator by means of another synchro-motor. The relative bearing dial is driven by a differential motor which receives two electric signals—antenna true bearing and "own ship's course." The rotor motion is the difference between these signals, or antenna relative bearing.

Similar arrangements can be used to transmit elevation data and to control the elevation angle of the antenna.

For fire control the antenna is directed at the target either manually, by automatic tracking, or by a combination of the two. In fire control the antenna true bearing and elevation are transmitted to the bearing and elevation indicators and also to the fire-control computer. The range reading can also be transmitted by synchros to the computer, when range is read from a dial as described in SEC. **9**.

When completely automatic tracking is used, the antenna rapidly scans a small angular region on all sides of its mean position. The asymmetry of the target signals on either side of the mean position develops an error signal that actuates a servo-system and causes it to reduce the error in mean position.

In so-called aided-tracking the fire-control computer sends back to the antenna servo-system an "error" signal that causes the antenna to follow the course of the target as predicted by the computer. The antenna is made to follow deviations of the target from the predicted course either by manual control or by automatic tracking.

The authors wish to thank the members of the staff of the Radar School for the opportunity of discussing with them the content of these articles, and especially to thank Professors W. H. Radford and H. J. Zimmerman, who have read the manuscript. We are also indebted to Mr. R. D. Spence for his help on Sec. 5. For assistance with the drawings we owe much to Miss Helen Cabot and her staff.

Internationality in the Names of Scientific Concepts: A Proposed International Photometric System

Parry Moon
Massachusetts Institute of Technology, Cambridge, Massachusetts

AND

Domina Eberle Spencer
Tufts College, Medford, Massachusetts

A PRECEDING paper[1] outlined a method of naming scientific concepts, such that both nonambiguity and internationality are achieved. The present paper gives an application of the general principles to a definite set of physical concepts. The concepts chosen are those dealing with radiometry, photometry and colorimetry.

Table I lists the usual names for the principal photometric concepts.[2] Evidently the internationality of these names is not notable. It is true that the word *illumination* is similar in English, Italian, Spanish and Portuguese; but in most other languages the corresponding word is unrecognizable except to the initiated. In French the concept is called *éclairement;* in German, *Beleuchtungsstärke;* and in Russian, *osveshchennost.* Words for *brightness* show even more variation. The *brilleance* of French, the *brillo* of Spanish, and the *brilho intrínseco* of Portuguese have some resemblance to the English word. But who could guess the precise meaning of the German *Leuchtdichte*, the Italian *splendore luminoso*, the Russian *i͡arkost*, and the Dutch *helderheid?*

The old photometric concepts are not unique in having noninternational names. Similar tabulations can be made for other sets of scientific concepts, with similar results. Whenever a word of common speech is forced to carry also a specialized technical meaning, the word will be translated into one or more terms for each language. In this way, the scientific world deliberately weights itself down with a heavy burden that is quite unnecessary and for which there is no excuse whatever.

Seventeenth-century and eighteenth-century science was almost free from the present trouble, since Latin was employed in most scientific writing. Near the beginning of the nineteenth century, however, Nationalism decreed that we were not to have Science but we must have German science and British science and French science. This movement has grown ever since,[3] until the situation has become almost intolerable.[4]

[1] P. Moon and D. E. Spencer, "Internationality in the names of scientific concepts: A method of naming concepts," *Am. J. Physics* **14**, 285 (1946).

[2] Based on *International lighting vocabulary* (Commission Internationale de l'Eclairage, Teddington, England, 1938); L. D. Belkind, *Dictionary of lighting terminology* (Moscow, 1939).

[3] For example, *Researches of the Electrotechnical Laboratory* (Tokyo) were published in English up to 1931, when the growth of nationalism caused a shift to Japanese. It is doubtful if this change was of any appreciable value to Japanese scientists, yet the step almost completely eliminated the usefulness of the researches to the rest of the world.

[4] The difficulty of translating technical catalogs and advertising matter for foreign use is stressed by L. L. Sell in

TABLE I. Usual names of photometric concepts.

Symbol	Unit	English	French	Italian	Spanish	Portuguese	German	Dutch	Russian
F	lumen	luminous flux	flux lumineux	flusso luminoso	flujo luminoso	fluxo luminoso	Lichtstrom	lumineus vloed	svetovoĭ potok
D	lumen m^{-2}	luminous flux-density	densité de flux-lumineux	densità superficiale del flusso luminoso	densidad del flujo luminoso	densidade do fluxo luminoso	Flächen-lichstromdichte		plotnost potoka
E	lumen m^{-2}	illumination	éclairement	illuminamento	iluminación	iluminação	Beleuchtungsstärke	illuminatie	osveshchennost
L	lumen m^{-2}	radiance, luminosity	radiance	luminosità	radiancia	luminosidade	spezifische Lichtausstrahlung		svetīmost
Q	lumen-sec	quantity of light	quantité de lumière	quantità di luce	cantidad de luz	quantidade de luz	Lichtmenge		kolīchīestvo sveta
U	lumen sec m^{-2}	exposure	excitation	eccitazione	exposición	exposição	Belichtung		ekspozītsseīa
I	candle	luminous intensity	intensité lumineuse	intensità luminosa	intensidad luminosa	intensidade luminoso	Lichstärke	lichtsterkte	sīla sveta
B	candle m^{-2}	brightness	brillance	splendore luminoso	brillo	brilho intrínseco	Leuchtdichte	helderheid	īarkost

The realization of the need of some sort of practical remedy is indicated by the appearance of a host of specialized technical dictionaries. Unfortunately, they are surprisingly ineffective, as anyone who has used them will testify. To return to Latin for scientific publications is quite impracticable, though there is a chance of employing an international auxiliary language such as Esperanto, Ido or Iala. Lacking this rather visionary solution of the difficulty, we can still do something by choosing future scientific words in such a way that they will automatically become universal. A study of present words will also show many cases where nomenclature can be made international.

In choosing names for devices and concepts, we need merely take words that are not in common use, say by selecting them from classical Greek. Experience shows that these words will not be translated but will be accepted as international.[1] Also, since these new scientific words do not have multiple meanings, we have simultaneously achieved a one-to-one correspondence between idea and word, which will do much to eliminate confusion within a given language. The following sections show how these principles can be applied to radiometry, photometry and colorimetry.

1. Proposed Active Concepts

Before considering the names, one must decide on the concepts that are to be used. In photometrics there is no fixed number of concepts. Many combinations of fundamental quantities are possible, and the investigator must try to strike a happy medium between the extreme of a separate symbol for every possible combination and the other extreme of insufficient concepts to meet everyday needs. As indicated previously,[5] we have found that six active photometric con-

his *Syllabus for the formation of the professional polyglot technician* (International Dictionary Co., New York, 1945). See also the preface to his *English-Spanish comprehensive technical dictionary* (McGraw-Hill Book Co., New York, 1944).

. The trouble caused by the lack of a one-to-one correspondence between words and scientific ideas is treated by A. S. Litvinenko in *Dictionary of radio terminology in the English, German, French, and Russian languages* (Moscow, 1937), preface.

[5] P. Moon, "A system of photometric concepts," *J. Opt. Soc. Am.* **32**, 348 (1942); "The names of physical concepts," *Am. J. Physics* **10**, 134 (1942).

TABLE II. Active radiometric concepts.

Symbol	Dimensions	mks Unit	Common name of concept	Proposed English name
F_r	$[P]$	watt	radiant power	radiant pharos
D_r	$[PL_t^{-2}]$	watt m^{-2}	radiant power density	radiant pharosage
Q_r	$[PT]$	watt sec	radiant energy	radiant phos
U_r	$[PTL_t^{-2}]$	watt sec m^{-2}	radiant energy density	radiant phosage
H_r	$[PL_t^{-4}L_r^2]$	herschel	——	radiant helios
G_r	$[PL_t^{-4}L_r]$	herschel m^{-1}	——	radiant heliosent
$J(\lambda)$	$[PL_t^{-2}\Lambda^{-1}]$	watt m^{-2} micron^{-1}	spectral radiant power density	phengosage

cepts are sufficient for all necessary calculations. In this system, the antiquated concept of *intensity*, or "candlepower," is scrapped. The general concept of *helios* replaces the idea of *brightness*, which was restricted to surface sources.[6] Also the general concept of *pharosage*, or *flux density*, replaces the two ideas of *illumination* and *luminosity*.

In considering photometric concepts, one finds it advantageous to treat also the related system of radiometric concepts. It is desirable to consider the dimensionality of the units associated with these concepts. Instead of using the dimensions $[L]$, $[M]$ and $[T]$, we find it convenient[5] to employ $[L]$, $[P]$ and $[T]$, where $[P]$ is power. In radiometry and photometry, solid angles are of frequent occurrence, and it is found that confusion results from considering them as dimensionless. We shall use two dimensions of length $[L_r]$ and $[L_t]$—radial and tangential—so that the dimensions of a plane angle are $[L_tL_r^{-1}]$, and those of a solid angle are $[L_t^2L_r^{-2}]$. Furthermore, *wavelength* should be distinguished from *length;* the dimension of wavelength will be denoted by $[\Lambda]$.

The proposed concepts in the active *radiometric* group are listed in Table II, and their dimensions and mks units are given. As the basic concept of all radiometry, photometry and colorimetry, one may take *radiant power* F_r. It has dimension $[P]$ and is measured in watts. The other concepts of Table II are associated with F_r by simple spacial or temporal relations. Thus D_r is radiant power per unit area. The quantities Q_r and U_r are obtained from F_r and D_r by multiplying by time:

$$Q_r = \int F_r dt, \quad U_r = \int D_r dt. \tag{1}$$

It is desirable also to have a concept that depends on the solid angle ω. Such a concept is defined by the equation[7]

$$H_r = \pi \lim_{\omega \to 0} D_r/\omega. \tag{2}$$

If D_r is measured in watts per square meter, and ω is in steradians, then H_r is said to be in *herschels*. The quantity G_r is merely the rate of change of H_r with respect to distance. The final concept, $J(\lambda)$ of Table II, is employed in the specification of continuous spectral distributions. If ΔD_r is the radiant power per unit area in a wavelength band of width $\Delta\lambda$, then the ratio of these two quantities is called $J(\lambda)$:

$$J(\lambda) = \lim_{\Delta\lambda \to 0} \Delta D_r/\Delta\lambda. \tag{3}$$

Spectral distribution curves for sources[8] are plots of $J(\lambda)$ *versus* λ.

The corresponding *photometric* quantities are obtained from the radiometric by weighting the latter with respect to vision. An additional dimension must be introduced, the most convenient being $[F]$, which denotes the photometric counterpart of radiant power. Table III presents the proposed concepts of the active photometric group. The dimensional correspondence with the

TABLE III. Active photometric concepts.

Symbol	Dimensions	Unit	Common name of concept	Proposed English name
F_l	$[F]$	lumen	luminous flux	luminous pharos
D_l	$[FL_t^{-2}]$	lumen m^{-2}	luminous flux density	luminous pharosage
Q_l	$[FT]$	lumen sec	quantity of light	luminous phos
U_l	$[FTL_t^{-2}]$	lumen sec m^{-2}	exposure	luminous phosage
H_l	$[FL_t^{-4}L_r^2]$	blondel	brightness	luminous helios
G_l	$[FL_t^{-4}L_r]$	blondel m^{-1}	——	luminous heliosent

[6] "Illuminating engineering nomenclature and photometric standards," *Illum. Eng.* **36**, 815 (1941).

[7] P. Moon and D. E. Spencer, "Photometrics in general physics," *Am. J. Physics* **11**, 200 (1943); "Brightness and helios," *Illum. Eng.* **39**, 507 (1944).

[8] P. Moon and D. E. Spencer, "Approximations to Planckian distributions," *J. App. Physics* **17**, 506 (1946); "Analytic representation of spectroradiometric curves," to be published.

TABLE IV. Connectives.

Symbol	Dimensions	Unit	Common name of concept	Proposed English name
$\bar{y}(\lambda)$	$[FP^{-1}]$	young watt^{-1}	relative luminosity, visibility	lamprosity (spectral)
$v(\lambda)$	$[FP^{-1}]$	lumen watt^{-1}	luminosity, visibility	
$\bar{y}_t$	$[FP^{-1}]$	young watt^{-1}	luminous efficacy	lamprosity (total)
v_t	$[FP^{-1}]$	lumen watt^{-1}		
η	$[FP^{-1}]$	lumen watt^{-1}	over-all efficacy	actance
c	none	lumen young^{-1}	——	conversion factor

concepts of Table II is obvious. For each radiometric concept there is a photometric counterpart, whose dimensions are the same except for the substitution of $[F]$ in place of $[P]$. The one exception is $J(\lambda)$, which occurs only in the radiometric system.

It is advantageous to use the same basic symbols in the radiometric and photometric systems. This causes no confusion, since the subscript $_r$ or $_l$ will be employed whenever necessary. The names are the same in radiometric and photometric tables, but the adjectives *radiant* and *luminous* are employed if there is chance of ambiguity. In most practical work, attention is confined to either the radiometric or the photometric system. In such cases the subscripts and the adjectives are unnecessary and may be omitted. The use of the same symbols and names for corresponding concepts in the two systems effects a great simplification and allows the equations of geometrical photometrics[9] to be applied without change to radiometrics.

2. Connective Concepts

Relating the radiometric and photometric systems are the connective concepts of Table IV. By definition, the photometric quantity D_l is

$$D_l=\int_0^\infty v(\lambda)J(\lambda)d\lambda. \qquad (4)$$

Here $v(\lambda)$ is a weighting function that evaluates the physical quantity $J(\lambda)$ with respect to its visual effect. Evidently the dimensions of $v(\lambda)$ are $[FP^{-1}]$ and the quantity is expressed in *lumens per watt.*

[9] P. Moon and D. E. Spencer, "Analytic expressions in photometry and colorimetry," *J. Math. Physics* **25**, 111 (1946); "Analytic representation of trichromatic data," *J. Opt. Soc. Am.* **35**, 399 (1945).

The quantity $v(\lambda)$ applies to homogeneous radiation of wavelength λ. A similar quantity v_t applies to *any* radiation:

$$v_t=D_l/D_r. \qquad \text{(lumen watt}^{-1}\text{)} \qquad (5)$$

Evidently this quantity has the same dimensions and is exactly the same kind of thing as $\bar{y}(\lambda)$. Thus the two should have the same kind of symbol and the same kind of name. The symbols and names used in the past have not indicated this close similarity.

A third bridge between the purely physical quantities (Table II) and the photometric quantities (Table III) is denoted by η. Whereas $\bar{y}_t$ expresses the lumens per *radiated* watt, η expresses the lumens per watt *input* of a lamp. It thus denotes a property of a source rather than a property of a radiation.

The foregoing three connectives are all that are desirable. Unfortunately, however, the use of two systems of units for these quantities is firmly established. Instead of always employing $v(\lambda)$, the colorimetrist and illuminating engineer frequently use $\bar{y}(\lambda)$, which has a maximum value of approximately unity. This function has been standardized by international agreement and is tabulated in many places.[10] The function $v(\lambda)$ is exactly the same except for an experimentally determined conversion factor c:

$$v(\lambda)=c\bar{y}(\lambda). \qquad (6)$$

This conversion factor is adjusted to make the result of Eq. (4) come out in lumens per unit area as measured by visual comparison against a physical standard. The value of c is still somewhat in doubt, but appears to be about 647.8 when the "new" lumen of 1940 is employed.[11] The reciprocal of c has been erroneously called the "mechanical equivalent of light."

Corresponding to Eq. (4), the colorimetrists write

$$Y=\int_0^\infty \bar{y}(\lambda)J(\lambda)d\lambda. \qquad (4a)$$

[10] See, for instance, D. B. Judd, "Extension of the standard visibility function to intervals of one millimicron by third-difference osculatory interpolation," *J. Research National Bur. Standards* **6**, 465 (1931); A. C. Hardy, *Handbook of colorimetry* (Technology Press, Cambridge, Mass., 1936); or reference 6.

[11] D. L. MacAdam, "Note concerning the maximum luminous efficiency of radiant energy," *J. Opt. Soc. Am.* **35**, 615 (1935); also reference 9.

TABLE V. Proposed concepts and their names. Asterisk refers to noninternational names. Must be translated into each language.

Symbol	Dimensions	Unit	Name (English)
Active radiometric concepts			
F_r	$[P]$	watt	pharos (radiant)
D_r	$[PL_t^{-2}]$	watt m^{-2}	pharosage
Q_r	$[PT]$	watt sec	phos
U_r	$[PTL_t^{-2}]$	watt sec m^{-2}	phosage
H_r	$[PL_t^{-4}L_r^{2}]$	herschel	helios
G_r	$[PL_t^{-4}L_r]$	herschel m^{-1}	heliosent
$J(\lambda)$	$[PL_t^{-2}\Lambda^{-1}]$	watt m^{-2} micron^{-1}	phengosage
Connectives			
$v(\lambda)$	$[FP^{-1}]$	lumen watt^{-1}	lamprosity (spectral)
v_t	$[FP^{-1}]$	lumen watt^{-1}	lamprosity (total)
η	$[FP^{-1}]$	lumen watt^{-1}	actance
Ratios expressing passive properties			
$\rho(\lambda)$	None	Numeric	reflectance (spectral)
$\tau(\lambda)$			transmittance (spectral)
ρ_r			reflectance (radiant)
τ_r			transmittance (radiant)
α_r			absorptance (radiant)
κ_r	$[L_r^{-1}]$	m^{-1}	afantivity (radiant)
$\epsilon(\lambda)$	None	Numeric	stilbance (spectral)
ϵ_t			stilbance (total)

Symbol	Dimensions	Unit	Name (English)
Active photometric concepts			
F_l	$[F]$	lumen, young	pharos (luminous)
D_l	$[FL_t^{-2}]$	lumen m^{-2}	pharosage
Q_l	$[FT]$	lumen sec	phos
U_l	$[FTL_t^{-2}]$	lumen sec m^{-2}	phosage
H_l	$[FL_t^{-4}L_r^{2}]$	blondel	helios
G_l	$FL_t^{-4}L_r]$	blondel m^{-1}	heliosent
Ratios expressing passive properties			
ρ_l	None	Numeric	reflectance (luminous)
τ_l			transmittance
α_l			absorptance
κ_l	$[L_r^{-1}]$	m^{-1}	afantivity
Dimensionless ratios occurring in room lighting			
ξ	None	Numeric	interflectance
k_u			logance
k_r			coefficient of utilization*
k_m			domance
T			luance
			delos

Symbol	Dimensions	Unit	Name
Colorimetric terms (CIE system)			
$\bar{x}(\lambda)$	$[F_xP^{-1}]$	könig watt^{-1}	trichromatic weighting functions*
$\bar{y}(\lambda)$	$[FP^{-1}]$	young watt^{-1}	
$\bar{z}(\lambda)$	$[F_zP^{-1}]$	priest watt^{-1}	
X	$[F_xL_t^{-2}]$	könig m^{-2}	trichromatic coordinates*
Y	$[FL_t^{-2}]$	young m^{-2}	
Z	$[F_zL_t^{-2}]$	priest m^{-2}	
(X, Y, Z)			*trichos*, a hypernumber having the coordinates X, Y, Z
x, y, z			homogeneous coordinates*
(x, y)			*chros*, a hypernumber having the coordinates x and y
λ_d	$[\Lambda]$	micron	dominant wavelength*
λ_c	$[\Lambda]$	micron	complementary wavelength
p			*akratos*

This is really another expression for D_l, but expressed in a different unit,[12] the *young per square meter*. Corresponding to Eq. (5), we have

$$\bar{y}_t = Y/D_r. \qquad \text{(young watt}^{-1}\text{)} \quad (5a)$$

This quantity is frequently used by illuminating engineers. It is called luminous efficacy, or luminous "efficiency," and is usually said to be a dimensionless fraction. Obviously, it cannot be dimensionless, since Y and D_r do not have the same dimensions. It is dimensionally identical with v_t, Eq. (5), but is expressed in different units. This brief discussion of the connective concepts may give an idea of the unnecessary complexity and confusion now existing.

3. Names of Concepts

The proposed concepts and names are assembled in Table V. First come the active concepts, the radiometric and photometric ones being listed side by side to show their similarity. All of the recommended names are taken from classical Greek, with the endings arranged in accordance with Table V of reference 1. Connecting the two systems are the quantities called *lamprosity* and *actance*. The quantities $v(\lambda)$ and v_t are determined primarily by the peculiarities of the photochemical substances in the retina, and hence they are classed as -ITY concepts. They are expressed in lumens per *radiated* watt. The quantity η, on the other hand, is expressed in lumens per watt *input* to a lamp. It therefore includes all the losses in the lamp and is generally regarded as a property of this device. Hence it is considered as an -ANCE concept.

Besides these fundamental concepts, there are a number of important ratios expressing reflecting, transmitting, absorbing and radiating properties. The official names, *reflection factor*, *transmission factor* and *absorption factor* are being

[12] P. Moon and D. E. Spencer, "Units in the trichromatic system," *J. Opt. Soc. Am.* **36**, 120, 306 (1946).

rapidly supplanted by the shorter *reflectance*, *transmittance* and *absorptance*. The -ANCE ending is correct, since the quantity depends not only on the material but also on the thickness and on the condition of the surface. In other words, the concept refers to a passive property of an entity rather than of a substance.

Fundamental in the specification of reflecting properties of a surface is the *spectral reflectance* $\rho(\lambda)$. It is a purely physical quantity and specifies the fraction of the incident radiant power that is reflected at each wave-length λ. Also convenient are the total reflectances ρ_r and ρ_l, which give the fraction of the integrated pharosage that is reflected:

$$\rho_r = \frac{\int_0^\infty \rho(\lambda)J(\lambda)d\lambda}{\int_0^\infty J(\lambda)d\lambda}, \tag{7}$$

$$\rho_l = \frac{\int_0^\infty \rho(\lambda)v(\lambda)J(\lambda)d\lambda}{\int_0^\infty v(\lambda)J(\lambda)d\lambda}. \tag{8}$$

The three transmittances are analogous to the reflectances.

For the exponential factor κ in Bouguer's relation,

$$H = H_0 e^{-\kappa l}, \tag{9}$$

we propose the name *afantivity* from the Greek word ἄφαντος meaning "made invisible," "blotted out." This is a true property of a material rather than a device, and therefore the ending -ITY is used. As in the case of reflectance, three different afantivities seem desirable: $\kappa(\lambda)$ refers to the behavior of homogeneous radiation of wave-length λ, κ_r considers the integrated effect for radiant power, and κ_l is the corresponding photometric quantity.

The quantity ϵ is usually called the *emissivity* of a radiator. By definition,

$$\epsilon(\lambda) = \frac{[J(\lambda)]_{\text{radiator}}}{[J(\lambda)]_{\text{blackbody}}}, \tag{10}$$

$$\epsilon_l = \frac{[D_r]_{\text{radiator}}}{[D_r]_{\text{blackbody}}}. \tag{11}$$

TABLE VI. Greek words used in the proposed system.

Greek	Transliteration	Original meaning
φῶς	phos	light
φάρος	pharos	a lighthouse
φέγγος	phengos	light, splendor
ἥλιος	helios	the sun
λαμπρός	lampros	bright, radiant
ἄφαντος	afantos	made invisible, blotted out
ἀκτίς	actis	a ray, beam of light
στιλβόω	stilbo	gleam of bright surfaces
λόγος	logos	ratio
δῆλος	delos	visible
δόμος	domos	house, room, chamber
χρώς	chros	color
λούω	louo	wash, bathe
ἄκρατος	akratos	unmixed, pure

Obviously, the ending -ITY is incorrect since ϵ refers to a complete device. We suggest the word *stilbance* from the Greek στιλβόω. All the Greek words employed in making the proposed system are listed in Table VI.

A set of dimensionless ratios occurring in lighting design is included in Table V. *Interflectance f* is defined as pharos to the illuminated plane ("working plane") divided by pharos from the luminares. It is a property of the room, depending particularly on the multiple reflections of light among the various surfaces, and is properly given an -ANCE ending. *Logance* is the ratio of the luminous output of the luminaire to the luminous output of the lamps that are within the luminaire[13]; it is usually called *efficiency*, but a better plan is to reserve the latter word for ratios of power or energy. The shape of the room is expressed by *domance*, from the Greek word δόμος (room). The reduction in average pharosage with age, caused by blackening of the lamps, depreciation of the paint and collection of dust, is called *luance*; this word may be employed in place of *maintenance factor*[13] or *depreciation factor*. Finally, it is desirable to have a word that will express visual acuity under the given conditions in terms of best possible visual acuity; *delos* is used for this purpose.[14]

The final part of Table V deals with colorimetric terms. The only new words here are *chros*, *trichros* and *akratos*.[15] *Chros* is suggested

[13] P. Moon and D. E. Spencer, "Maintenance factors," *Illum. Eng.* **41**, 211 (1946).

[14] P. Moon and D. E. Spencer, "Visual data applied to lighting design," *J. Opt. Soc. Am.* **34**, 605 (1944); "The visual effect of non-uniform surrounds," *J. Opt. Soc. Am.* **35**, 233 (1945).

[15] The accent is on the first syllable.

as a substitute for the usual term *chromaticity*,[16] in which the ending -ITY is employed in a non-international sense. Chros (x, y) is a hypernumber having the two coordinates x and y. It indicates a quantitative property of the radiation, one that is closely associated with color "quality." Since the concept represents a property of radiation, it is an -OS concept and should not have a name ending in -ITY.

Trichros (X, Y, Z) is an analogous quantity in color three-space. It is a hypernumber having the three coordinates X, Y and Z. It may be considered as an affine vector in color space, one end of the vector being at the origin of coordinates and the other at the point (X, Y, Z). This concept provides an outstanding example of the inadequacy of the old method of naming concepts. The concept is usually called *color*. But there are so many other meanings for the word *color* that its use in this highly specialized meaning is found to be almost impossible: there is not a mere chance that the reader will misunderstand; there is almost complete certainty that he will misunderstand. Hence the need for a new word, such as *trichros!*

The third new word in this part of Table V is *akratos*. It represents what is usually called *purity*.[16] Like *chromaticity*, this concept specifies a characteristic of radiation, not a property of a material. Thus the customary -ITY ending is inadvisable.

4. Names in Various Languages

A previous paper[1] has shown that words obtained by the method there outlined are international so that they can be assimilated by all the languages that are being used extensively in scientific work. Table V shows how the proposed words appear in English. It is possible that these words could be incorporated without change into all European languages, but the words will seem more natural if they conform somewhat to the orthography of the ethnic language. Table VII shows how the proposed words would appear in eight natural languages.

This table includes all the new words of Table

TABLE VII. Proposed international names for radiometric and photometric concepts.

Symbol	International	English	French	Italian	Spanish	Portuguese	German	Dutch	Russian
	Active concepts								
F	faroso	pharos	pharosse	faroso	faroso	faroso	Pharos	pharos	faros
D	farosago	pharosage	pharossage	farosaggio	farosago	farosago	Pharosag	pharosag	farosazh
Q	foso	phos	phosse	foso	foso	foso	Phos	phos	fos
U	fosago	phosage	phossage	fosaggio	fosago	fosago	Phosag	phosag	fosazh
H	helioso	helios	héliosse	elioso	helioso	helioso	Helios	helios	ėlīos
G	heliosento	heliosent	héliossent	eliosento	heliosento	heliosento	Heliosent	heliosent	ėlīosent
$J(\lambda)$	fengosago	phengosage	phengossage	fengosaggio	fengosago	fengosago	Phengosag	phengosag	fengosazh
	Connectives								
$\bar{y}(\lambda), \bar{y}_t$	lamprosito	lamprosity	lamprosité	lamprosità	lamprosidad	lamprosidade	Lamprosität	amprositeit	lamprosītī
η	aktanco	actance	actance	attanza	actancia	actância	Aktanz	aktantie	aktant͡s
	Ratios expressing passive properties								
ρ	reflektanco	reflectance	réflectance	riflettenza	reflectancia	reflectância	Reflektanz	reflektantie	rīflektant͡s
τ	transmitanco	transmittance	transmittance	transmittenza	transmitancia	transmitância	Transmittanz	transmittantie	transmītant͡s
α	absorbtanco	absorptance	absorptance	assorbenza	absortancia	absortância	Absorptanz	absorptantie	absorbtant͡s
κ	afantito	afantivity	afantivité	afantità	afantividad	afantividade	Afantivität	afantiviteit	afantīvītī
ϵ	stilbanco	stilbance	stilbance	stilbanza	estilbancia	estilbância	Stilbanz	stilbantie	stīlbant͡s
	Dimensionless ratios used in room lighting								
f	interflektanco	interflectance	interflectance	interflettenza	interflectancia	interflectância	Interflektanz	interflektantie	īnterflektant͡s
g	loganco	logance	logance	loganza	logancia	logância	Loganz	logantie	logant͡s
k_r	domanco	domance	domance	domanza	domancia	domância	Domanz	domantie	domant͡s
k_m	luanco	luance	luance	luanza	luancia	luância	Luanz	luantie	luant͡s
T	deloso	delos	delosse	deloso	deloso	deloso	Delos	delos	delos
	Colorimetric terms								
(X, Y, Z)	trikroso	trichros	trichrosse	tricroso	tricroso	tricroso	Trikros	trikros	trīkros
(x, y)	kroso	chros	chrosse	croso	croso	croso	Kros	kros	kros
p	akratoso	akratos	akratosse	acratoso	acratoso	acratoso	Akratos	akratos	akratos

[16] Report of the Colorimetry Committee, *J. Opt. Soc. Am.* **34**, 246 (1944).

TABLE VIII. Alternative names for ρ, τ, α.

Symbol	Previous word	Alternative
ρ	reflectance	anaphosance
τ	transmittance	diaphosance
α	absorptance	esophosance
f	interflectance	metaphosance

Greek prefixes used in the above names: δια- =dia- =through; ἀνα- =ana- =back; ἐσω- =eso- =into; μετα- =meta- =among.

V. In some cases, however, it seems inadvisable to introduce international terms. Such words as *time*, *length*, *wavelength* and *dominant wavelength* may be left in their original forms in each language. It seems best to leave unchanged *trichromatic weighting functions*, *trichromatic coordinates* and *homogeneous coordinates*. The other words are listed in Table VII. Evidently the suggested words are easily recognizable in all eight languages, and the list can be extended to cover other languages.

5. Alternative Names for ρ, τ, α

The only troublesome part of Table VII is the part that employs the names *reflectance*, *transmittance* and *absorptance*. The Latin roots for these words are so firmly established that they may be regarded as essentially international, and for that reason they were retained in Table VII. Yet the terms actually used in engineering practice have always been of the form *facteur d'absorption d'un corps* or *fattore di assorbimento di un corpo*, and the results of an attempt to reduce such expressions to single words with standardized endings were not very satisfactory. The difficulty becomes acute if one attempts to employ also the -ITY ending. For instance, *assorbtità* is a poor word, made even more repulsive to any Italian by the fact that it is not a frankly foreign word but contains elements of his own language. Another possibility, *assorbimentità* is probably worse. Even the *absorptivity* of English is not good.

To see what could be done with Greek, we made up the alternative set of names given in Table VIII. The common part *phos* (φῶς = light) shows that the concepts deal with radiation, while the Greek prefixes indicate whether the light is reflected, transmitted, absorbed or interflected. The ending -ANCE or -ITY indicates whether the concept refers to a filter, room or other entity, or to a material. These words do not belong to any nation or language and could be accepted internationally. The older terms are so firmly established, however, that most people would undoubtedly hate to replace them by the unfamiliar words of Table VIII.

It is interesting to note that the active concepts of Table V are not the only ones that are named by the proposed system. We have limited ourselves to the smallest number of concepts that will allow all calculations to be made conveniently. But if anyone wishes to employ more active concepts, he is at liberty to do so and the names are automatically provided for him. Among the possibilities are *pharosent* (lumens per meter of fluorescent tubing), *pharosum* (lumens per cubic meter of a luminous vapor), radiant *phosum* (energy per unit volume), *phengos* (watts per micron), *phengosent* (watts per meter, per micron), *phengosum* (watts per cubic meter, per micron). Evidently the proposed system of nomenclature has considerable flexibility. In the interest of simplicity, however, it seems best in most cases to limit ourselves to the concepts listed in Table V.

The present paper has shown that the established names of photometric concepts are widely different in different languages and therefore unsatisfactory as regards internationality. In eliminating this difficulty, one can eliminate simultaneously much of the ambiguity and complexity associated with the old system.[17] The number of concepts can be reduced, the confusion between popular and scientific meanings for the same word can be banished, and a truly international set of names can be developed. Names are suggested for radiometric, photometric and colorimetric concepts, and the form of these words is given in nine languages.

[17] An interesting study of internationality in science is given in E. Wüster, *Internationale Sprachnormung in der Technik* (Berlin, 1931).

UNTIL they have learned to express themselves, scientists will continue to be wallflowers at the world's quickstep.—D. W. HILL, *The impact and value of science.*

Reproduction of Prints, Drawings and Paintings of Interest in the History of Physics

28. The First Hydrogen Balloon

E. C. Watson
California Institute of Technology, Pasadena 4, California

Jacques Alexandre César Charles (1746–1823), the French physicist who discovered the gas law which bears his name and who was the first to fill and to ascend in a hydrogen balloon, was born on November 12, 1746. It is appropriate to commemorate this bicentennial by reproducing Charles' portrait together with several of the many interesting contemporary prints that celebrated the success of his pioneer experiments in aerostatics.

The first public ascent of a large scale balloon took place at Annonay, France, on June 5, 1783. The balloon used was of the hot-air type invented and constructed by the Montgolfier brothers. The demonstration was so successful that the Académie des Sciences in Paris was stimulated to raise money for similar experiments and work with hydrogen was undertaken by Charles. The ascent of the first hydrogen balloon was described by Benjamin Franklin in a letter to Joseph Banks, President of the Royal Society in London, as follows:[1]

Passy, Aug. 30, 1783

Sir,

On Wednesday the 27th instant, the new aerostatic experiment, invented by Messrs. Mongolfier of Annonay was repeated by Mr. Charles; Professor of Experimental Philosophy at Paris.

A hollow globe 12 feet diameter was formed of what is called in England oiled silk, here Taffetas *gommée*, the silk being impregnated with a solution of gum-elastic in lintseed oil, as is said. The parts were sewed together while wet with the gum, and some of it was afterwards passed over the seams, to render it as tight as possible.

It was afterwards filled with the inflammable air that is produced by pouring oil of vitriol upon filings of iron, when it was found to have a tendency upwards so strong as to be capable of lifting a weight of 39 pounds, exclusive of its own weight which was 25 lb. and the weight of the air contain'd.

It was brought early in the morning to the *Champ de Mars*[2] a field in which reviews are sometimes made, lying between the Military School and the river. There it was held down by a cord, till 5 in the afternoon, when it was to be let loose. Care was taken before the hour to replace what portion had been lost of the inflammable air, or of its force, by injecting more.

It is supposed that not less than 50,000 people were assembled to see the experiment. The Champ de Mars being surrounded by multitudes, and vast numbers on the opposite side of the river.

At five o clock notice was given to the spectators by the firing of two cannons, that the cord was about to be cut. And presently the globe was seen to rise, and that as fast as a body of 12 feet diameter with a force of only 39 pounds, could be suppos'd to move the resisting air out of its way. There was some wind, but not very strong. A little rain had wet it, so that it shone, and made an agreeable appearance. It diminished in apparent magnitude as it rose, till it enter'd the clouds, when it seem'd to me scarce bigger than an orange, and soon after became invisible, the clouds concealing it.

The multitude separated, all well satisfied & much delighted with the success of the experiment, and amusing one another with discourses of the various uses it may possibly be apply'd to, among which many were very extravagant. But possibly it may pave the way to some discoveries in natural philosophy of which at present we have no conception.

A note secur'd from the weather had been affix'd to the globe, signifying the time & place of its departure, and praying those who might happen to find it, to send an account of its state to certain persons at Paris. No news was heard of it till the next day, when information was receiv'd, that it fell a little after 6 oclock at Gonesse, a place about 4 leagues distance; and that it was rent open, and some say had ice in it. It is suppos'd to have burst by the elasticity of the contain'd air when no longer compress'd by so heavy an atmosphere.

One of 38 feet diameter is preparing by M. Mongolfier himself at the expence of the Academy, which is to go up in a few days. I am told it is constructed of linen & paper, and is to be filled with a different air, not yet made public, but cheaper than that produc'd by the oil of vitriol of which 200 Paris pints were consum'd in filling the other.

It is said that for some days after its being fill'd, the ball was found to lose an eighth part of its force of levity in 24 hours: Whether this was from imperfection in the tightness of the ball, or a change in the nature of the air, experiments may easily discover.

I thought it my duty, Sir, to send an early account of this extraordinary fact, to the Society which does me the honour to reckon me among its members; and I will

[1] This letter, which is preserved in the library of the University of Pennsylvania, is reproduced in *The ingenious Dr. Franklin*, by N. G. Goodman (Philadelphia, 1931).

[2] Where the Eiffel Tower now stands.—E. C. W.

PLATE 1. Jacques Alexandre César Charles (1746–1823). [From a print published by Berthoud, Paris.]

endeavor to make it more perfect, as I receive farther information.

With great respect, I am, Sir,

B. FRANKLIN

P.S.

I just now learn, that some observers say, the ball was 150 seconds in rising, from the cutting of the cord till hid in the clouds; that its height was then about 500 toises,[3] but, mov'd out of the perpendicular by the wind, it had made a slant so as to form a triangle, whose base on the earth was about 200 toises. It is said the country people who saw it fall were frightened, conceiv'd from its bounding a little when it touch'd the ground, that there was some living animal in it, and attack'd it with stones and knives, so that it was much mangled; but it is now brought to town & will be repaired.

The great one of M. Mongolfier, is to go up as is said, from Versailles, in about 8 or 10 days. It is not a globe but of different form, more convenient for penetrating the air. It contains 50,000 cubic feet, and is supposed to have a force of levity equal to 1500 pounds weight. A philosopher here, M. Pilatre de Rozier, has seriously apply'd to the Academy for leave to go up with it, in order to make some experiments. He was complimented on his zeal and courage for the promotion of science, but advis'd to wait till the management of these balls was made by experience more certain & safe. They say the filling of it in M. Mongolfier's way will not cost more than half a crown. One is talk'd of to be 110 feet diameter. Several gentlemen have ordered small ones to be made for their amusement; one has ordered four of 15 feet diameter each; I know not with what purpose; but such is the present enthusiasm for promoting & improving this discovery, that probably we shall soon make considerable progress in the art of constructing and using the machines:—

Among the pleasantries conversation produces on this subject, some suppose flying to be now invented, and that since men may be supported in the air, nothing is wanted but some light handy instruments to give and direct motion. Some think progressive motion on the earth may be advanc'd by it, and that a running footman or a horse slung & suspended under such a globe so as to leave no more of weight pressing the earth with their feet, than perhaps 8 or 10 pounds, might with a fair wind run in a straight line across countries as fast as that wind, and over hedges, ditches, & even waters. It has been even fancied that in time people will keep such globes anchored in the air, to which by pullies they may draw up game to be preserved in the cool, & water to be frozen when ice is wanted. And that to get money, it will be contrived to give people an extensive view of the country, by running them upon an elbow chair a mile high for a guinea, etc., etc.

The first human beings to ascend in a balloon were JEAN FRANÇOIS PILÂTRE de ROZIER (1754–1785) and FRANÇOIS LAURENT, MARQUIS D'ARLANDES (1742–1809).[4] This ascent was made on November 21, 1783. Only ten days later, on December 1, 1783, the second ascent and the first in a hydrogen-filled balloon was made by CHARLES and one of the brothers ROBERT from the gardens of the Tuileries in Paris. The balloon, which was constructed by the brothers ROBERT, was made of lutestring coated with gum elastic and had a diameter of 27 ft. The car was suspended from a hoop surrounding the middle of the balloon and fastened to a net which covered the upper hemisphere. After ascending to a height of about 2000 ft and covering a distance of 27 mi in about 2 hr, CHARLES and ROBERT descended near the small town of Nesle, where ROBERT left the car and CHARLES reascended alone for a journey lasting a further 35 min, during which he reached a height estimated at 2 mi.

The pair carried thermometers, barometers and other "philosophical instruments" for the observation

[3] About 3200 ft; 1 toise≡6.395 ft.—E. C. W.

[4] See Reproduction 12 in this series, *Am. J. Physics* **8**, 249 (1940).

of as many new natural phenomena as might possibly be discovered in these hitherto uncharted regions. The voyage was completely successful and marked by no unexpected incidents. . . . Despite the elaborate collection of instruments, Professor Charles noted no new phenomena beyond the clearly predicted decrease in barometric pressure with height. He specifically noted only that the atmospheric temperature decreased so rapidly that in 10 minutes he passed "from the warmth of spring to the cold of winter."[5]

It may be of interest to quote also BENJAMIN FRANKLIN's report to JOSEPH BANKS upon this, the second aerial voyage to be made by man :[6]

Passy, Dec. 1, 1783

Dear Sir:—

In mine of yesterday I promised to give you an account of Messrs. Charles & Robert's experiment, which was to have been made this day, and at which I intended to be present. Being a little indisposed, and the air cool, and the ground damp, I declined going into the garden of the Tuileries, where the balloon was placed, not knowing how long I might be obliged to wait there before it was ready to depart, and chose to stay in my carriage near the statue of Louis XV., from whence I could well see it rise, and have an extensive view of the region of air through which, as the wind sat, it was likely to pass. The morning was foggy, but about one o'clock the air became tolerably clear, to the great satisfaction of the spectators, who were infinite, notice having been given of the intended experiment several days before in the papers, so that all Paris was out, either about the Tuileries, on the quays and bridges, in the fields, the streets, at the windows, or on the tops of houses, besides the inhabitants of all the towns and villages of the environs. Never before was a philosophical experiment so magnificently attended. Some guns were fired to give notice that the departure of the balloon was near, and a small one was discharged, which went to an amazing height, there being but little wind to make it deviate from its perpendicular course, and at length the sight of it was lost. Means were used, I am told, to prevent the great balloon's rising so high as might endanger its bursting. Several bags of sand were taken on board before the cord that held it down was cut, and the whole weight being then too much to be lifted, such a quantity was discharged as to permit its rising slowly. Thus it would sooner arrive at that region where it would be in equilibrio with the surrounding air, and by discharging more sand afterwards, it might go higher if desired. Between one and two o'clock, all eyes were gratified with seeing it rise majestically from among the trees, and ascend gradually above the buildings, a most beautiful spectacle. When it was about two hundred feet high, the brave adventurers held out and waved a little white pennant, on both sides of their car, to salute the spectators, who returned loud claps of applause. The

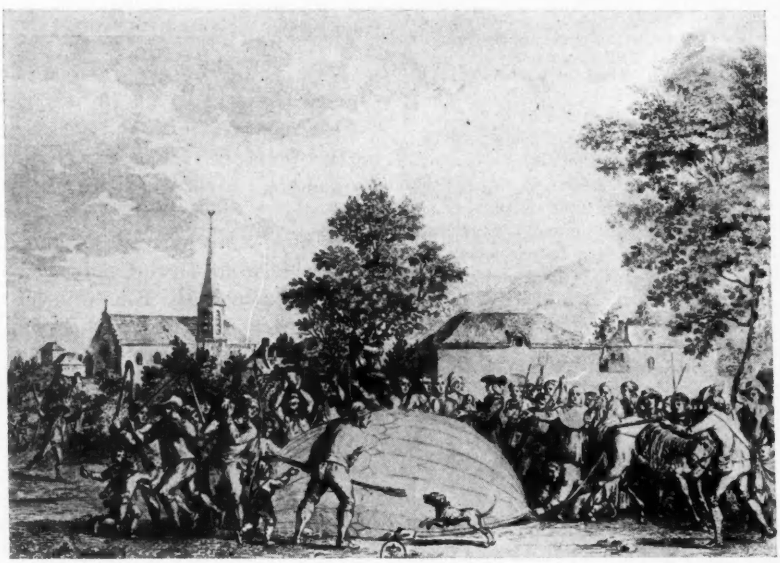


PLATE 2. The landing of the first hydrogen balloon, August 27, 1783, in the village of Gonesse. [From a contemporary French print.]

[5] F. A. Magoun and E. Hodgins, *A history of aircraft* (McGraw-Hill, 1931).
[6] This letter is also reproduced in reference 1.

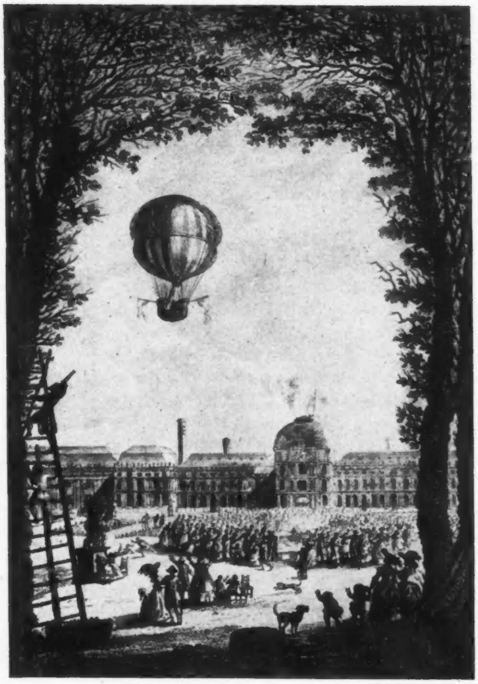


PLATE 3. The ascent of Charles and Robert from the Tuileries, December 1, 1783. [From an engraving by de Launay after de Lorimer, published by Vachez in Paris.]

wind was very little, so that the object though moving to the northward, continued long in view; and it was a great while before the admiring people began to disperse. The persons embarked were Mr. Charles, professor of experimental philosophy, and a zealous promoter of that science; and one of the Messieurs Robert, the very ingenious constructors of the machine. When it arrived at its height, which I suppose might be three or four hundred toises, it appeared to have only horizontal motion. I had a pocket-glass, with which I followed it, till I lost sight first of the men, then of the car, and when I last saw the balloon, it appeared no bigger than a walnut. I write this at seven in the evening. What became of them is not yet known here. I hope they descended by daylight, so as to see and avoid falling among trees or on houses, and that the experiment was completed without any mischievous accident, which the novelty of it and the want of experience might well occasion. I am the more anxious for the event, because I am not well informed of the means provided for letting themselves down, and the loss of these very ingenious men would not only be a discouragement to the progress of the art, but be a sensible loss to science and society.

I shall enclose one of the tickets of admission, on which the globe was represented, as originally intended, but is altered by the pen to show its real state when it went off. When the tickets were engraved the car was to have hung to the neck of the globe, as represented by a little drawing I have made in the corner.

I suppose it may have been an apprehension of danger in straining too much the balloon or tearing the silk, that induced the constructors to throw a net over it, fixed to a hoop which went around its middle, and to hang the car to that hoop.

Tuesday morning, December 2nd.—I am relieved from my anxiety by hearing that the adventurers descended well near L'Isle Adam before sunset. This place is near seven leagues from Paris. Had the wind blown fresh they might have gone much farther.

If I receive any further particulars of importance, I shall communicate them hereafter.

With great esteem, I am, dear sir, your most obedient and most humble servant,

B. FRANKLIN.

P. S. *Tuesday evening.*—Since writing the above I have received the printed paper and the manuscript containing some particulars of the experiment, which I enclose. I hear further that the travellers had perfect command of their carriage, descending as they pleased by letting some of the inflammable air escape, and rising again by discharging some sand; that they descended over a field so low as to talk with the labourers in passing, and mounted again to pass a hill. The little balloon falling at Vincennes shows that mounting higher it met with a current of air in a contrary direction, an observation that may be of use to future aerial voyagers.

Most of the features of modern balloons are due to CHARLES. Thus he was the first to use hydrogen successfully, and he invented the valve at the top of the balloon as well as the method of suspending the car which are still generally used.

CHARLES attained great fame during his lifetime, and many portraits of him exist. Plate 1 was made from a print published in Paris and reproduced by F. L. BRUEL in his monumental *Histoire Aéronautique par les Monuments, Peints, Sculptés, Dessinés et Gravés des Origines à 1830* (Paris, 1909). The vignette below the portrait-medallion depicts the enthusiastic scene when CHARLES and ROBERT landed at Nesle,

> . . . where the *procès verbal* was signed by, among others, the Duc de Chartres and a "Gentilhomme anglais," Mr. Farrer, who rushed up to Charles on his arrival with the explanation, "Moi, Charles, premier!" and was in such a state of excitement that he signed the *procès verbal* twice over in an almost illegible hand.[7]

[7] W. Lockwood Marsh, *Aeronautical prints and drawings* (London, 1924).

Plate 2 reproduces a contemporary print showing the frightened citizens of Gonesse attacking the first hydrogen balloon after its descent in the town. The scene and the reaction of the peasants is graphically described in the following quotation given in H. TURNOR'S *Astra Castra* (London, 1865):

> For on first sight it is supposed by many to have come from another world; many fly; others, more sensible, think it a monstrous bird. After it has alighted, there is yet motion of it from the gas it still contains. A small crowd gains courage from numbers, and for an hour approaches by gradual steps, hoping meanwhile the monster will take flight. At length one bolder than the rest takes his gun, stalks carefully to within shot, fires, witnesses the monster shrink, gives a shout of triumph and the crowd rushes in with flails and pitchforks. One tears what he thinks to be the skin, and causes a poisonous stench; again all retire. Shame, no doubt, now urges them on, and they tie the cause of alarm to a horse's tail, who gallops across the country, tearing it to shreds.

German and English copies of this print were published at Augsberg, Germany, and by JOHN WALLIS of Ludgate Street, London, and French copies in Paris.

The ascent of CHARLES and ROBERT from the Tuileries on December 1, 1783, produced a great profusion of prints.[8] Selected for reproduction here (Plate 3) is another[9] from the delightful series of DE LAUNAY after DE LORIMER, which was published by VACHEZ in both colored and uncolored states. These "are perhaps the best of all ballooning prints"[10]; several of them were used as illustrations in FAUJAS DE SAINT-FOND'S *Description des expériences de la machine aérostatique* (Paris 1783–4), which is "the chief contemporary authority for the details of the earlier ascents."[10]

Plate 4 shows CHARLES reascending alone after the signing of the *procès verbal* at Nesle. Plate 5 "is a highly imaginative picture of the triumphant

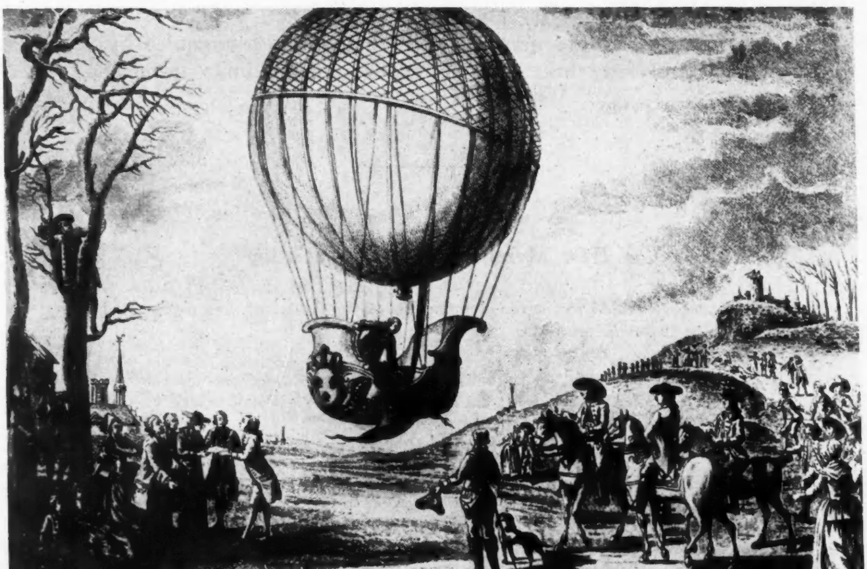


PLATE 4. Charles reascending alone after landing at Nesle, December 1, 1783. [From an engraving by Denis after Desrais published in Paris by Bassett.]

[8] The *Katalog der Historischen Abteilung der Ersten Internationalen Luftschiffahrts-Ausstellung (ILA) zu Frankfurt a. M. 1909*, by L. Liebmann and G. Wahl (Frankfurt, 1912), lists more than 40 without exhausting the list.
[9] See *Am. J. Physics* **8**, 250 (1940) for the other.
[10] Reference 7.

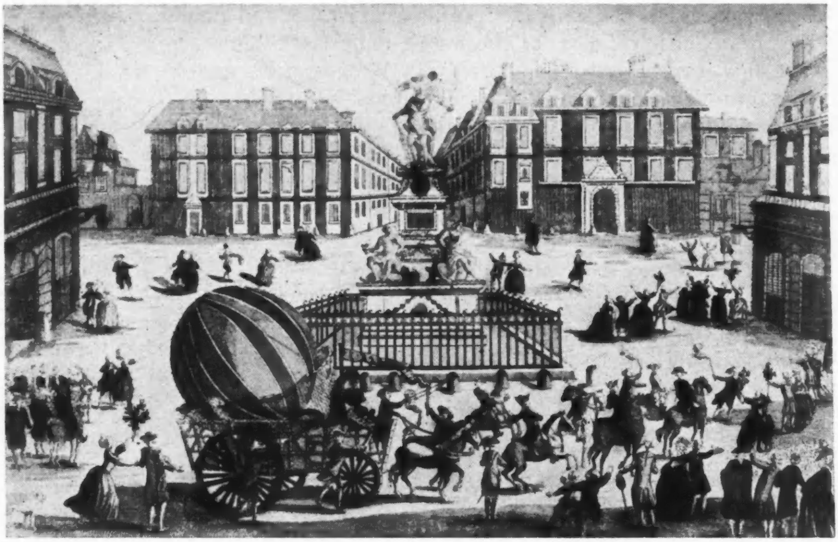


PLATE 5. The return of the Charles and Robert balloon to Paris, December 2, 1783. [From a color print published in Paris by Bassett.]

return of the balloon to Paris the next day, showing it apparently still fully inflated with hydrogen and surrounded by flaming torches in dangerous proximity to the gas."[10]

New Members of the Association

The following names have been added to the membership list of the American Association of Physics Teachers since the Directory of Members for 1946 was printed.

Bartlett, Col. B. W., Quarters 107, U. S. Military Academy, West Point, N. Y.
Blitzer, Leon, University of Arizona, Tucson, Ariz.
Bogardus, Theodore S., 141 Murphy St., Berea, Ohio.
Brown, Sheldon, 10455 Libson Lane, Los Angeles 24, Calif.
Buckle, George A., Polytechnic Institute of Brooklyn, Brooklyn 2, N. Y.
Buder, E. Edward, 4121a Potomac St., St. Louis, Mo.
Butler, Alfred B., 1822 D. St., Pullman, Wash.
Damas, Zion, Mont-sur-Marchiene, Belgium.
Fagan, Henry D., Xavier University, New Orleans 18, La.
Fowler, Richard G., Box 82, University of Oklahoma, Norman, Okla.
Green, Alex E. S., 703 Riddle Rd., Cincinnati 20, Ohio.
Lyman, Ernest M., 1009 S. Orchard St., Urbana, Ill.
McCarthy, Kathryn A., 128 Professor's Row, Medford 55, Mass.
McKay, Robert E., 605 Ridge St., Bowling Green, Ohio.
Madonna, Sister M. I.H.M., Immaculate Heart High School, 5515 Franklin Ave., Los Angeles 28, Calif.
Moe, Chesney R., San Diego State College, San Diego 5, Calif.
Olson, Manfred, 2112 E. Bradford Ave., State Teachers College, Milwaukee, Wis.
Oncken, William, Jr., Naval Ordnance Laboratory, Washington, D. C.
Payne, William T., University of Detroit, Detroit 21, Mich.
Rodgers, Thomas A., Box 123, State College, Miss.
Schneider, Edwin G., Stevens Institute of Technology, Hoboken, N. J.
Schultz, Frederick H. C., Box 303, University Station, Grand Forks, N. Dak.
Shannon, (Rev.) James I., S.J., St. Louis University, St. Louis, Mo.
Stollberg, Robert, Wabash College, Crawfordsville, Ind.
Suchy, Ray W., 2736 N. 10th St., Milwaukee, Wis.
Wade, Walter B., 813 11th St., SW, Birmingham, Ala.
Walker, Charles C., Orangeburg, S. C.
Woolsey, George, 4541 W. Ave., 40, Los Angeles 41, Calif.

Col. B. W. Bartlett has been a member of the Association for many years; his name was inadvertently omitted from the 1946 Directory of Members.

NOTES AND DISCUSSION

Lognormal Distribution

EVERETT F. COX
Washington, District of Columbia

IN his article on "Practical statistics for practical physicists,"[1] R. H. Bacon remarks that "it is not necessary to assume that the distribution is 'normal' in order to apply statistical methods . . . ." However, there is no doubt that practical physicists prefer to make calculations on normal distributions, for which the *Handbook of Chemistry and Physics* gives areas, ordinates, and second, third and fourth derivatives.

J. H. Gaddum[2] cites a number of cases of scientific observations whose *logarithms* are distributed normally. Examples include the values of house property, sizes of the foreheads of crabs, numbers of petals on buttercups and weights of female students. Gaddum concludes that if scientific observations which show variations that are large compared with themselves are converted to logarithms before estimating their mean or variance, the usual result is an increase in the accuracy and scope of conclusions drawn from them. Thus, much simpler statistical calculations may result if the experimenter uses semilog paper, or even log-log paper,[3] to plot frequencies of occurrence.[4]

[1] R. H. Bacon, *Am. J. Physics* **14**, 84, 198 (1946).
[2] J. H. Gaddum, *Nature* **156**, 463 (1945).
[3] P. Allen, S. C. Pearce and J. H. Gaddum, *Nature* **156**, 746 (1945).
[4] Logarithmic probability graph paper (No. 3128) is printed by the Codex Book Co., Norwood, Mass.

Uses for Electrically Charged Balloons in the Demonstration Lecture

PAUL ROOD
Western Michigan College of Education, Kalamazoo 45, Michigan

AMONG the many uses of electrically charged balloons in demonstrations are the following six:

(1) Two or three balloons suspended by threads make a clearly visible system to which to apply Coulomb's law.[1] The balloons are inflated with illuminating gas by means of a small motor-driven air pump, the illuminating gas being substituted for air at the intake. The inflation should be stopped just short of the point at which the buoyancy of the air will support the balloon. In the balloons used in these experiments a diameter of 22 cm was sufficient to give an apparent weight of about 2 gm. If two of these balloons are suspended by threads about 1.8 m long and then are rubbed over their entire surfaces with fur, they will stand apart about 60 cm (Fig. 1). The horizontal component of the weight of either balloon is then about 330 dynes. The Coulomb force of repulsion is equal in magnitude and opposite in direction to this, and the charge on each balloon is found to be 1100 statcoul. Since the potential at a point is the charge divided by the distance, the potential immediately beneath either balloon and due to the charges on both is 35,000 v.

(2) Electrostatic induction is shown by holding near the balloons a grounded metal rod with a large metal sphere at one end and a sharp point at the other. The balloons are attracted to the induced charge on the sphere but assume their original orientation when the sphere is removed.

(3) Discharge of electricity by points is demonstrated by holding the pointed end of the rod near a balloon (Fig. 1). Since most of the 35,000-v potential difference is concentrated near the point, the field causes ionization of the air, and positive ions are sprayed onto the balloons like paint from a spray gun. As a result the balloons quickly fall together.

(4) Ionization of the air by x-rays may be demonstrated. When a beam of x-rays is directed toward the balloons their charges are almost instantly neutralized.

(5) Ionization by flames is shown by placing a small flame under the balloons.

(6) A charged balloon is a convenient object to use for a variety of induction experiments. It is better than an ebonite rod because of the larger charge that can be stored on it. Suppose the sensitivity of the grid of a three-electrode vacuum tube is to be demonstrated. A short aerial is

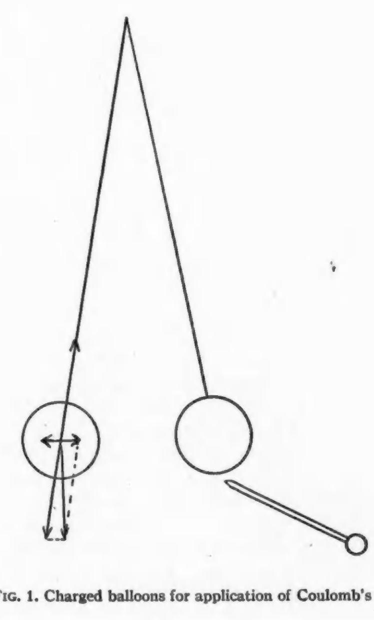


FIG. 1. Charged balloons for application of Coulomb's law.

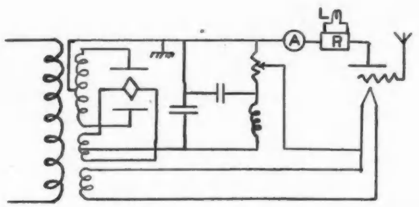


FIG. 2. Circuit for showing the sensitivity of the grid of a vacuum tube.

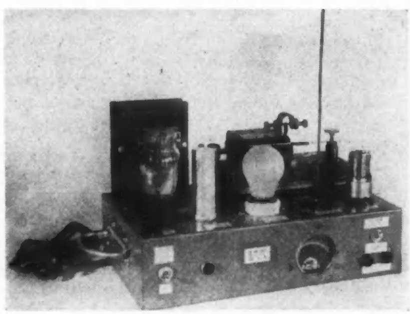


FIG. 3. Photograph of grid-controlled vacuum tube circuit.

attached to the grid of a tube in a circuit so arranged that the grid potential is determined by its position in the electric field between the filament and plate (Figs. 2 and 3). If a charged balloon is brought near to such an aerial, the potential of the grid is lowered and the plate current is suddenly cut to zero. This operates the relay R, and the signal lamp L goes out. But the grid, whose potential is now abnormally low, slowly returns to its usual condition by collecting positive ions from the residual gas in the tube. When it has reached its normal potential, the lamp again lights. The rate of returning to normal potential becomes higher as the tube gets hot because then more residual gas is driven from the walls.

If the balloon is now withdrawn the grid is left abnormally positive, and the plate current momentarily increases, as can be seen by watching the ammeter in the circuit. But electrons sufficient to neutralize the positive charge are quickly collected from the copious supply in the tube, and the plate current returns to its normal value.

It is instructive to repeat the experiment with a positively charged rod. When it is first brought near the aerial, the grid becomes abnormally positive but is quickly neutralized by electrons. If now the rod is removed, the electrons collected by the grid are trapped on it and reduce its potential so much that there will be no plate current, as is indicated by the extinction of the signal lamp. It will stay off until the negatively charged grid has collected enough positive ions to neutralize these electrons. As in the preceding experiment this action becomes more rapid as the tube gets hot.

[1] P. Rood, *Am. J. Physics* **8**, 320 (1940).

A Model to Show the Perfect Focusing of a Parabolic Mirror

F. R. HIRSH, JR.
University of Southern California, Los Angeles 7, California

THE writer has previously described two models[1] for spherical mirrors, demonstrating spherical aberration. The next logical step is to demonstrate the perfect focusing of a parabolic mirror.

The present model (Fig. 1) was made of scrap lumber, as in the previous two cases. It is $2\times1\frac{1}{2}$ ft in size. The parabolic track is made of plywood, the inner layer of wood being cut through at right angles to the radius of curvature to facilitate bending. The black arrows on the white "rays" indicate the direction of light travel. As

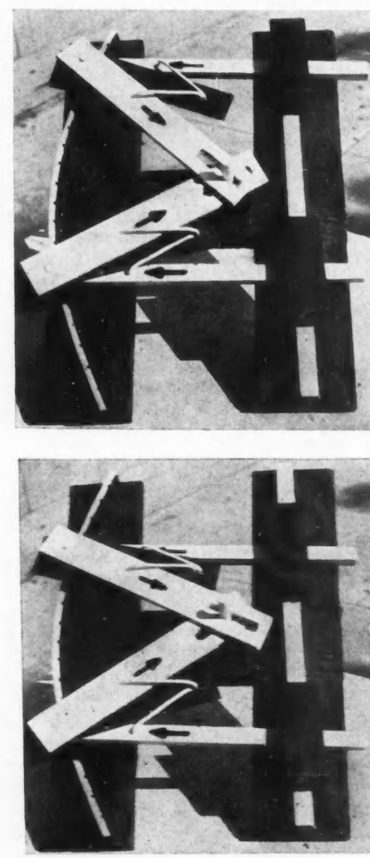


FIG. 1. Model of a parabolic mirror.

before, a parallelogram device keeps the angle of incidence equal to the angle of reflection. It will be seen that the reflected rays always pass through the principal focus (white peg).

[1] F. R. Hirsh, Jr., *Am. J. Physics* **13**, 267 (1945); **14**, 66 (1946).

NEWS OF SECTIONS

Oregon Section

THE forty-second meeting of the Oregon Section of the American Association of Physics Teachers was held on April 27, 1946 at the N. W. Electrodevelopment Laboratory of the U. S. Bureau of Mines, Albany, Oregon. The following program was presented.

Modern prospecting. S. H. Lorain, *Albany, Oregon.*

A simplified method for comparing densities in half-tone prints. F. S. Minshall, *University of Oregon.*

The material of interstellar space. E. B. Ebbighausen, *University of Oregon.*

A discussion of the Young People's meeting at Reed College. A. A. Knowlton, *Reed College.*

Under the guidance of Dr. B. A. Rogers, those attending the meeting took a trip through the Electrodevelopment Laboratory.

Officers elected for 1946–47 are: *President*, E. T. Broan; *Secretary-Treasurer*, W. R. Varner.

The next meeting will be at Lewis and Clark College, Portland.

W. R. VARNER, *Secretary*

Southern California Section

On June 1, 1946 was held the second annual competitive physics test for high school seniors, sponsored by the Southern California Section of the American Association of Physics Teachers. The test was given at Occidental College, Santa Barbara State College, San Diego State College and Redlands University. All high school seniors whose names were submitted and certified by a high school teacher or principal were eligible. One hundred twenty-four students competed. The examination consisted of a matching test involving definitions of basic physical quantities and problems in mechanics, electricity, heat, light and sound. Honor scholarships having a value of $300 to $800, offered by Occidental College, Pomona College and Redlands University, and a physics award of the value of $58 offered by the University of California at Los Angeles were available as prizes, to be selected by the winners on the basis of their rank. In addition, certificates of award were given to the first ten contestants in order of their scores.

The committee in charge of the examination consisted of V. L. Bollman, Chairman, Occidental College; C. R. Coony, S.J., Loyola University of Los Angeles; F. K. Hurd, University of California at Los Angeles; A. Nye, University of Southern California.

R. W. MCHENRY, *Secretary*

Announcement of Oersted Medalist for 1946

THE Committee on Awards of the American Association of Physics Teachers announces the selection of Dr. Duane Roller as the 1946 Oersted Medalist. The actual award will be made as usual at the annual meeting, which this year is at Columbia University, January 30–31 and February 1, 1947. Heretofore the identity of the recipient has not been disclosed until the actual time of presentation of the award. Henceforth, it will be announced in advance.

Doctor Roller needs no introduction to readers of the *American Journal of Physics*. He has been the editor of this journal from its first issue in 1933 under the name of "The American Physics Teacher." The 76 subsequent issues constitute a contribution to physics that any man could be proud to cite as his life work. The Oersted medal is awarded "for notable contributions to the teaching of physics." The *Journal* is outstandingly the most notable contribution to the teaching of physics at the college and university levels since the subject was introduced into our educational system, and the Committee takes great pride in the selection of Doctor Roller as this year's recipient of the award.

The citation and address of acceptance will appear in full in a forthcoming issue of the *Journal*. The present note is merely the preliminary announcement of this year's award.

LLOYD W. TAYLOR
For the Committee on Awards

DIGEST OF PERIODICAL LITERATURE

Academic Upheaval

THE atom bomb has brought about certain changes in one of the large, ivy-clad, midwestern state universities. The university has always had a College of Liberal Arts and Sciences, but the most influential faculty members were in the College of Agriculture—the professors of Agronomy, Animal Husbandry, Beekeeping, etc., who were, as a class, men of large frame, large families, and unlimited confidence. They took command of faculty meetings, knowing that if there was a showdown they would be backed by the state legislators of the rural sectors. Their progress across the campus on a sunny day was impressive, almost gladiatorial.

Into this atmosphere, twelve years ago, came William Ames, B.S. (Harvard), Ph.D. (Oxon.), Professor of Physics. Ames was a slight man, with a progressive myopia which obliged him to wear thick-lensed spectacles. His wife, Clara, affected low-heeled, sensible shoes and was pretty in a dull way. Her remarks were never quoted. At faculty teas she was not asked to pour. She had no children, a circumstance which was emphasized by the bountiful fecundity of the other campus wives. In an effort to overcome this misfortune, she and the Professor read thick compendiums on genes, mutations of species, anthropology, and the breeding habits of aboriginal tribes.

As a student at Oxford, Ames had sat at the feet of the great Lord Rutherford and had learned the true and unrelenting nature of the laws that govern our unseen world. His classes in Advanced Physics, over the years, had few takers. His requests for apparatus for research were shouted down in faculty meetings by the Professor of Animal Husbandry and the Professor of Meat Cutting, who had their eyes on new cream separators and slicing machines, or by the Professor of Agronomy, who wanted a new tractor with four-wheel drive and radio. Professor Ames was a quiet man, a lamb among the rams, and resigned to his colorless career. Secretly, he admired the physique and the bullying manner of the Professor of Meat Cutting.

The summer of 1945 was idling by at the university, Germany had been defeated, the Army Specialized Training Program classes were drowsing in the heat of the plains. On August 5th, in one microsecond, the light of ten thousand suns broke over Hiroshima and an ugly mushroom cloud poked its head sixty thousand feet in the air. Nine days later the war was over. There was jubilation on the campus for a week, and then a hush set in as the papers began to relate in detail the work of the scientists, the long-haired scientists, the *physicists*, who held the secret of world military power in their formulas.

At the first faculty meeting of the school year, in October, the full professors were gathered in a large knot near the door, talking among themselves, waiting for the president to call them to their places. Simultaneously with the banging of his gavel, a hush fell over the group, and there was gentle shuffling and a parting of the mass as Professor Ames walked into the room and over to his accustomed place. He looked neither to the right nor to the left, and his small frame was as erect as a Prussian colonel general's at a color mount. As the meeting progressed, he asked for the floor, which he was given immediately and with sweeping deference. He stated that his classes in Advanced Physics were up to his expectations. With an enrollment of two hundred and fifty students, it was necessary to get an amount of new apparatus—a mass spectrometer, a small cyclotron, and some other electrical equipment. All in all, the Physics Department would need, roughly, a million two hundred thousand dollars. He turned around and, glaring at the assembly, asked if there were any objections from other departments. The Professor of Meat Cutting rose like a cobra listening to a flute and moved that the appropriation be made. It was seconded and passed. Professor Ames then turned on his heel, said, "Good day, gentlemen," and left the hall by the center aisle, erect, looking neither to the right nor to the left.

On May 13, 1946, the Professor's wife, Clara, gave birth to twin sons. Since then, Ames has been observed from time to time in local saloons, his foot on the brass rail, lecturing the barkeepers in loud tones.—JOHN T. COX, JR., *The New Yorker*, Sept. 7, 1946, p. 29. Permission of *The New Yorker;* copyright 1946, The F-R. Publishing Corporation.

Postdoctoral Fellowship of Sigma Delta Epsilon

Women with the equivalent of a Ph.D. degree, carrying on research in the mathematical, physical or biological sciences, who need financial assistance and give evidence of high ability and promise are eligible for this fellowship. The stipend is $1500. During the term of appointment the appointee must devote the major part of her time to the approved research project, and not engage in other work for remuneration unless such work shall have received the written approval in advance of the award of the fellowship.

Application blanks may be secured from Dr. Louise S. McDowell, 28 Dover Road, Wellesley 81, Massachusetts. Applications for the year 1947–1948 should be submitted before February 1, 1947.

Author Index to Volume 14

In this index are listed the names of authors and the titles of their articles and notes. Abstracts of papers read at meetings and of articles appearing under "Digest of Periodical Literature" in the journal are not listed; however, they are indexed in the Analytic Subject Index.

Analytic Subject Index to Volume 14

The titles of articles are disregarded, the entries being based on analyses of the contents of the original articles. (D) designates a digest of an article or an abstract of a paper read at a meeting; (T) designates title only.

To facilitate reference to any desired subject, the index is divided into sections arranged alphabetically. The titles of these sections and of cross references to them are as follows:

6

—

nal

nd
nd

per

(T)
ers,

ning

L.

ncer

om-

309

. G.

—54,